

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

На правах рукопису
УДК 004.942:519.216.3

До захисту допущено
В. о. завідувача кафедри ММСА

О.Л.Тимошук

«__» _____ 2018 р.

Магістерська дисертація

на здобуття ступеня магістра за спеціальністю 124 Системний аналіз
на тему: «Система підтримки прийняття рішень для впровадження
маркетингових стратегій у банку»

Виконала:

студентка II курсу, групи КА-72 мп

Гребенюк Вікторія Миколаївна

Керівник: Професор кафедри ММСА

д.т.н, професор, Бідюк П.І.

Рецензент: Професор кафедри ЗЗІ

ФТІ НТУУ "КПІ ім. І. Сікорського"

д.т.н, професор, О.Є. Архипов

Засвідчую, що у цій магістерській дисертації
немає запозичень з праць інших авторів
без відповідних посилань

Студент _____

Київ
2018

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ
СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Рівень вищої освіти — другий (магістерський)

Спеціальність (спеціалізація) — 124 «Системний аналіз» («Системний аналіз і управління»)

ЗАТВЕРДЖУЮ

В. о. завідувача кафедри ММСА

О. Л. Тимощук

«___» _____ 2018 р.

ЗАВДАННЯ

на магістерську дисертацію студентці Гребенюк Вікторії Миколаївні

1. Тема дисертації: «Система підтримки прийняття рішень для впровадження маркетингових стратегій у банку», науковий керівник дисертації Бідюк Петро Іванович, доктор технічних наук, професор, затверджені наказом по університету від «07» листопада 2018 р. № 4121-с

2. Термін подання студентом дисертації:

3. Об'єкт дослідження: позичальники кредитів, представлені статистичними даними з вибраними характеристиками

4. Предмет дослідження: принципи формування маркетингової стратегії в банку, математичні моделі, методи і критерії оцінювання адекватності скорингових моделей та методи побудови скорингової карти

5. Перелік завдань, які потрібно розробити:

- 1) Огляд технічної літератури за темою роботи;
- 2) Вивчення сучасних процесів при впровадженні маркетингової стратегії у банку;
- 3) Дослідження актуальності теми дослідження;
- 4) Вибір методів для побудови моделі та набору вхідних даних;

- 5) Опис архітектури системи підтримки прийняття рішень;
- 6) Детальний огляд всіх обраних методів;
- 7) Вибір параметрів тренування для кожного з них та критеріїв їх ефективності;
- 8) Проведення аналізу ринкових можливостей запуску стартап-проекту.

6. Орієнтовний перелік графічного (ілюстративного) матеріалу:

- 1) Схема стратегії маркетингу в банку
- 2) Архітектура системи підтримки прийняття рішень
- 3) Вигляд моделі скорингу в програмі Azure ML Studio
- 4) Знімки екрану CRM системи для впровадження маркетингової стратегії

7. Орієнтовний перелік публікацій:

- (1) Система підтримки прийняття рішень для впровадження маркетингових стратегій у банку

8. Дата видачі завдання:

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Отримання завдання на дипломну роботу	07.09.2018	
2	Огляд технічної літератури за темою	07.09.2018 - 01.10.2018	
3	Дослідження актуальності вибраної теми	02.10.2018 - 07.10.2018	
4	Вибір методів для побудови моделі	07.10.2018 – 17.10.2018	
5	Пошук наборів вхідних даних	17.10.2018 – 19.10.2018	
6	Опис архітектури системи підтримки прийняття рішень	19.10.2018 – 23.10.2018	
7	Детальний огляд всіх обраних методів	24.10.2018 – 29.10.2018	
8	вибір параметрів тренування для кожного з них та критеріїв їх ефективності	30.10.2018 – 02.11.2018	
9	Проведення аналізу ринкових можливостей запуску стартап-проекту	02.11.2018 – 19.11.2018	
10	Написання слайдів для доповіді	20.11.2018 – 26.11.2018	

Студент
Науковий керівник дисертації

В.М. Гребенюк
П.І. Бідюк

РЕФЕРАТ

Магістерська дисертація: 103 с., 31 рис., 5 табл., 16 джерел.

Об'єкт дослідження – набір статистичних даних щодо наданих банком споживчих кредитів фізичним особам.

Предмет дослідження – принцип формування маркетингової стратегії в банку, математичні моделі, методи і критерії оцінювання адекватності скорингових моделей.

Методи дослідження – методи інтелектуального аналізу даних, регресійний аналіз, статистичні методи аналізу даних, методи побудови скорингових моделей.

Метою роботи є розробка системи підтримки прийняття рішень для впровадження маркетингової стратегії у банку за допомогою оцінювання та прогнозування кредитоспроможності населення у режимі реального часу використовуючи скоринговий метод. В роботі представлений новий підхід до процесу прийняття рішень щодо реалізації маркетингової кампанії у банку на основі побудови скорингових моделей, розглянуто та проаналізовано методи логістичної регресії, дерев рішень та нейронних мереж. Було проаналізовано результати моделювання та оцінювання задля обґрунтованого вибору найкращої моделі для оцінки кредитоспроможності клієнтів.

Розроблено архітектуру системи підтримки прийняття рішень результатом якої є підвищення лояльності та довіри клієнтів банку та зниження витрат на маркетинг. Для побудови та аналізу математичних моделей використано пакет програмного забезпечення Microsoft Azure Machine Learning Studio.

БАНКІВСЬКИЙ МАРКЕТИНГ, МАРКЕТИНГОВА СТРАТЕГІЯ У БАНКУ, КРЕДИТНИЙ СКОРИНГ, СКОРИНГОВА МОДЕЛЬ, СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ, МАРКЕТИНГОВА КАМΠΑНІЯ У БАНКУ

ABSTRACT

The topic: Decision support system for implementing marketing strategies in bank.

Master's thesis: 103 p., 28 fig., 31 tab., 16 sources.

Object of the study – loans borrowers represented by the statistics of selected characteristics.

Subject of the research – principle of the formation of marketing strategies in the bank, mathematical models and methods that describe heteroscedastic processes, estimation and analysis of the quality of forecasts, and the estimation models of market risks.

Methods of research – data mining, neural networks, regression modelling, decision trees, classification methods, scoring models development.

The aim is to develop a decision support system for implementing marketing strategies in bank via evaluating and predicting the creditworthiness of the clients using scoring techniques.

In this paper, new approaches for decision making process in banking marketing based on constructing scoring models, reviewed logistics methods, decision trees and neural networks is presented. It analyzes the results of modeling and assessment studies for choosing the best model for assessing the creditworthiness of customers. Result of the development lies in increasing the loyalty and trust of the bank's clients and reducing marketing costs. For the construction and analysis of mathematical models was used the software package Microsoft Azure Machine Learning Studio. An example of implementing a marketing campaign was chosen by the Microsoft Dynamics 365 CPM system.

**BANKING MARKETING, MARKETING STRATEGY IN THE BANK,
CREDIT SCORING, SCORING MODELS, DECISION SUPPORT SYSTEM,
MARKETING CAMPAIGN IN BANK**

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ І ТЕРМІНІВ	10
ВСТУП.....	11
РОЗДІЛ 1 ОСОБЛИВОСТІ ВПРОВАДЖЕННЯ МАРКЕТИНГОВОЇ СТРАТЕГІЇ В БАНКІВСЬКІЙ УСТАНОВІ.....	13
1.1 Сутність, функції та принципи банківського маркетингу	13
1.2 Сутність та функції банківської стратегії.....	17
1.3 Сучасний підхід до формування банківської стратегії для кожного цільового ринку	27
1.4 Скоринг	29
1.5 Скорингові моделі	29
1.6 Системи підтримки прийняття рішень	30
1.7 Етапи проектування СППР	32
Висновки до розділу	36
РОЗДІЛ 2 ВИБІР І ОПИС МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЛЯ СКОРИНГУ	37
2.1 Різновиди машинного навчання	37
2.2 Параметри, що необхідні для визначення алгоритму	40
2.3 Вибір алгоритмів машинного навчання.....	41
2.4 Лінійні регресійні моделі оцінювання кредитоспроможності клієнтів за скоринговим методом	43
2.5 Алгоритм дерев рішень оцінювання кредитоспроможності клієнтів	52
Висновки до розділу	55
РОЗДІЛ 3 РОЗРОБКА СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ НА ОСНОВІ ПОБУДОВИ СКОРИНГОВИХ МОДЕЛЕЙ.....	57
3.1. Формулювання вимог до СППР	57
3.2 Розробка архітектури та функціональної схеми програми.....	58
3.3 Вибір інструментальної платформи для виконання обчислювальних експериментів	61

3.4 Використання отриманих результатів моделі для впровадження маркетингової стратегії в банку	67
- Висновки до розділу	68
РОЗДІЛ 4 ПОБУДОВА СКОРИНГОВИХ МОДЕЛЕЙ ТА АНАЛІЗ РЕЗУЛЬТАТІВ.....	70
4.1 Трансформація, очистка та опис отриманих даних.....	70
4.2 Проведення експериментів для визначення найкращої скорингової моделі	80
4.2.1 Логістична регресія	80
4.2.2 Древа рішень	84
4.3 Порівняння отриманих результатів	87
Висновки до розділу	89
РОЗДІЛ 5 АНАЛІЗ РИНОВИХ МОЖЛИВОСТЕЙ СТАРТАП-ПРОЕКТУ	90
5.1 Опис ідеї проекту	90
5.2 Технологічний аудит ідеї проекту.....	91
5.3 Аналіз ринкових можливостей запуску стартап-проекту.....	92
5.4 Розроблення ринкової стратегії проекту	99
5.5 Розроблення маркетингової програми стартап-проекту.....	102
Висновки до розділу	106
ВИСНОВКИ ПО РОБОТІ ТА ПЕРСПЕКТИВИ ДЛЯ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ.....	108
ПЕРЕЛІК ПОСИЛАНЬ	110

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ І ТЕРМІНІВ

ІАД – інтелектуальний аналіз даних;

ОПР – особа що приймає рішення;

ПП – програмний продукт;

ПЗ – програмне забезпечення;

ММП – метод максимальної правдоподібності;

СППР – система підтримки прийняття рішень;

AIC – Akaike info criterion (інформаційний критерій Акайке);

DSS - Decision Support System.

ВСТУП

Робота присвячена системному вирішенню питання впровадження маркетингових стратегій у банківських установах. Банківська діяльність є ефективною, коли створений і успішно функціонує маркетинговий підрозділ, що забезпечує попит на ринку виробленого банківського продукту або наданих ним послуг, а задоволення певних потреб клієнтів, завдяки використанню даного продукту або послуги, приносить прибуток.

Сьогодні в Україні спостерігається тенденція зростання конкуренції між комерційними банками. На території нашої країни здійснюють свою діяльність близько вісімдесяти банків, які пропонують клієнтам схожий набір послуг. Тому керівництву банку необхідно своєчасно прийняти рішення щодо вдосконалення своєї діяльності, а одним з ключових завдань є пошук шляхів і методів для створення стійких переваг в умовах все більш конкурентоспроможних фінансових ринків. Роль маркетингу в рамках цього завдання не викликає сумнівів, особливо в сфері якості банківських послуг.

Якість досягається шляхом винесення потреб клієнтів на перший план та надання їм спеціального обслуговування. Історично так склалось, що використання маркетингових елементів у банку розпочалось пізніше, ніж у інших сферах підприємницької діяльності. Даний лаг пояснюється початковою орієнтацією банків на юридичних осіб.

Застосування інформаційних технологій може зробити цей спосіб функціонування набагато продуктивнішим, а банк конкурентоспроможним на ринку фінансових послуг. Наразі у кожному банку існує база даних, яка містить важливу інформацію про клієнтів. Правильна обробка таких даних, побудова моделі, аналіз та формування відповідних висновків може надати банку конкурентну перевагу, а споживачам- якісне обслуговування.

Відомими сучасними теоретиками і практиками в області інтелектуального аналізу даних є професори Ю.П. Зайченко, В.М. Подладчиков, Н.Д. Панкратова, Джонатан Н. Крук, Лін С. Томас, Девід Дж. Хенд, Л.М. Любчик, доктор Елізабет Мейз, дослідники Наїм Сіддікі, Девід Б. Едельман.

Метою роботи є розробка системи підтримки прийняття рішень для впровадження маркетингової стратегії у банку за допомогою оцінювання та прогнозування кредитоспроможності населення у режимі реального часу використовуючи скоринговий метод.

Цілями даної системи підтримки прийняття рішень є :

- підвищення лояльності та довіри клієнтів банку;
- зниження витрат на маркетингові кампанії;
- зменшення кількості відмов у наданні кредитів;
- зменшення кількості заборгованості;
- збільшення операційної ефективності.

РОЗДІЛ 1 ОСОБЛИВОСТІ ВПРОВАДЖЕННЯ МАРКЕТИНГОВОЇ СТРАТЕГІЇ В БАНКІВСЬКІЙ УСТАНОВІ

1.1 Сутність, функції та принципи банківського маркетингу

Одним з основних елементів просування банку на ринку будь-яких послуг в умовах ринкової економіки є маркетинг.

Маркетинг (від англ. Market- ринок) – представлення ринкової теорії управління.

Основна особливість банківського маркетингу- орієнтація принципів управління в банківських установах на кредитно-фінансові системи. Об'єктом маркетингових задач є гроші.

Існує два види концепцій маркетингу:

- традиційна;
- сучасна.

Традиційна концепція маркетингу полягає в наданні освоєних банківських послуг інструментами організації збуту для отримання прибутків за рахунок збільшення об'єму збуту. Банківська установа розвиває технологічні та облікові аспекти вдосконалення послуг.

Сучасна концепція орієнтується на потреби клієнтів і за рахунок роботи маркетингового відділу досягає збільшення прибутку за рахунок більш повного задоволення потреби клієнта. Ця орієнтація на зусилля і можливості банківської установи на визначенні реальних і потенційних потреб більшості суб'єктів економічних відносин і здійснення пошуків способу їх найбільшого вдоволення, виходячи з фінансового, кадрового, організаційного, технологічного, законодавчого та інших обмежень.

Наразі для більшості вітчизняних банків характерна традиційна, хоча вже з'являються банки, які використовують сучасну концепцію маркетингу.

Для того щоб глибше зрозуміти особливості маркетингових стратегій в банківській установі розглянемо визначення таких основних понять: банківського продукту і послуги.

Банківський продукт - це набір певних банківських послуг. *Банківська послуга* це набір певних операцій, що включені до банківського продукту. Банківський продукт не є матеріальним, хоча для того щоб його створити необхідно пройти певний виробничий процес (аналіз ринку, проектування продукту, створення, тестування та розповсюдження). Особливостями банківського продукту порівняно з іншими(матеріальними) є:

- не має матеріального втілення;
- довгостроковість відносин споживача та виробника(банку);
- нерозграничність процесу створення та застосування банківського продукту;

Банківський продукт та банківська послуга схожі за функціями, які виконують, а саме задоволення потреби клієнта та збільшення прибутку.

Слід зауважити, що всі банківські операції контролюються державою, а це в свою чергу накладає певні обмеження на властивості банківських продуктів. Тому банківські установи в певній мірі більш обмежені у виробництві продуктів ніж інші підприємства.

Активні банківські операції (прикладом може бути надання кредиту) несуть високий ризик для банку, адже неправильна оцінка позичальника може призвести до суттєвих збитків банківської установи, що знецінить її репутацію на ринку. Тому на відділ маркетингу у банку покладено більше функцій та відповідальності.

Досвід функціонування банківського маркетингу в діяльності світових банківських установ дає можливість зробити висновки, що маркетинг у банківській установі є особливою економічною категорією.

Банківський маркетинг (як економічна категорія) – це відносини між суб'єктами кредитно-фінансових відносин — банками, фінансовими посередниками, споживачами (фізичними та юр.особами) з приводу задоволення потреб і вирішення проблем клієнтів банку в банківських продуктах, що призводить до економічного успіху банків і клієнтів та приносить користь суспільству в цілому [1].

Окремої уваги заслуговує визначення запропоноване Л.Ф. Романенко у монографічному дослідженні "Банківський маркетинг", яке вийшло друком у 2001 році. Автором вперше зроблена спроба консолідувати два головних підходи до визначення суті банківського маркетингу як філософії банківської діяльності і як способу здійснення підприємницької політики банку. У результаті автором зроблено висновок, що банківський маркетинг - це "філософія, стратегія й тактика банку, що спрямовані на ефективне задоволення потреб, запитів і сподівань, вирішення проблем існуючих і потенційних банківських клієнтів шляхом успішної реалізації на ринку банківських продуктів, які враховують тенденції розвитку ринку та приносять користь суспільству завдяки управлінню активами і пасивами, прибутками і витратами, ліквідністю банку, рівнем ризику його операцій"[2].

Сутність маркетингу в банківській установі можна передати через опис основних принципів, а саме:

- принципи задоволення потреби клієнтів;
- принципи орієнтації на довгостроковій перспективі;
- принципи досягнення кінцевого результату.

Принцип вдоволення потреби клієнта полягає у поєднанні дослідницьких та виробничих заходів банківської установи для вдоволення найбільш повного набору потреб клієнта, швидкої реакції на зміни попиту на продукти банківської установи та передбачення незадоволених потреб клієнта та втілення їх в нових продуктах банку. Виявлення таких потреб сприяє виходу на нові ринки та сегменти фінансового ринку.

Принципи орієнтації на довгостроковій перспективі (пов'язаний з функцією управління маркетингом) полягає у спрямованості на довгострокові результати та, як наслідок прибуток у майбутній перспективі, за допомогою дослідження ринку фінансово-кредитних послуг, прогнозуванню попиту клієнтів та впровадженню нових послуг банку.

Принцип досягнення кінцевого результату, а саме збільшення прибутковості банківської установи за рахунок виходу чи створення нових ринків та покриття нових сегментів споживачів послуг банку. Виходячи з цього принципу в банку має бути розроблена програма дій та завдань на визначений часовий проміжок для досягнення планового обсягу прибутку чи обсягу активів.

Для досягнення визначених цілей необхідно виконати сукупність завдань, а саме:

- дотримання визначеного обсягу послуг, що пропонується банком визначеному переліку клієнтів, у визначений проміжок часу;
- збільшення планового обсягу користування банківськими послугами;
- завоювання певної частини ринку одною з типів послуг банку в заплановані терміни;
- планове зростання рівня задоволення користувачів послуг якістю продуктів.

Також, для виконання стратегічних цілей, обов'язковими є такі задачі:

- зміна чи впровадження внутрішньої корпоративної культури банку з вираженою клієнтоорієнтованістю;
- забезпечення підтримки професійного розвитку та навчання персоналу;
- впровадження прогресивних інформаційних технологій;
- формування відносин партнерів між банківською установою та клієнтом.

До ключових функцій маркетингу в банківській установі відносять :

- визначення реального і можливого ринкового попиту і вимог клієнтів до продуктів, що надає банківська установа за допомогою дослідження актуального стану фінансових ринків та прогнозування їх розвитку в майбутньому;
- оптимізація наданих та активне створення нових послуг та продуктів банку, що найбільш повно задовольняють вимоги споживачів;
- визначення оптимальної кількості представлених на ринку продуктів банку, враховуючи реальні можливості банківської установи та ринковий попит;
- створення збутової та цінової стратегії банківської установи;
- створення та підтримка внутрішнього та зовнішнього іміджу банківської установи, як надійної, що надає якісні банківські послуги та задовольняє найактуальніші потреб клієнта;

1.2 Сутність та функції банківської стратегії

Маркетингова стратегія банку – постановка довгострокових цілей банку та опис заходів для задоволення потреб клієнтів та отримання прибутку у майбутньому.

Маркетингова стратегія являється концепцією управління діяльністю банку, що має бути направлена на вивчення ринку та його кон'юнктури, визначення потреб потенційних та існуючих клієнтів та створення банківських продуктів, що будуть задовольняти ці потреби.

Основними цілями маркетингової стратегії банку є:

- збільшення клієнтської бази;
- збільшення якості та об'єму пропонованих банківських продуктів;
- збільшення чистого прибутку банківської установи

- формування та стимулювання попиту на банківські продукти

Для досягнення визначених цілей банк виконує наступні задачі:

- налагодження партнерських відносин з клієнтом;
- маркетингові комунікації по просуванню банківського продукту;
- стимулювання робітників банку продажу банківських продуктів.

Стратегія складається з двох самостійних категорій: стратегії та планування. Під стратегією розуміємо мистецтво управління діяльністю комерційним банком, яке поєднує у собі теорію та практику поведінки з клієнтами. Планування – це інструмент збільшення клієнтської бази.

Таким чином маркетингова стратегія в банку розділяється на дві частини:

- формулювання, розробка, постановка цілей та задач банку, а також підтримка відповідності теперішнім цілям та можливостям комерційного банку;
- розробка програми дій банку(реалізація вибраної стратегії банку).

Суть основних задач розробки маркетингової стратегії банку полягає в максимальному задоволенні вимог клієнтів в об'ємі, властивостях та якості банківських продуктів. Таким чином створюються найкращі умови для формування довгострокових ділових партнерських відносин; забезпечення рентабельної роботи банку у постійній невизначеності фінансового ринку; забезпечення ліквідності банку. Ефективна діяльність банку має на увазі наявність добре опрацьованого комплексу різних маркетингових технологій, що сприяють своєчасно реагувати на зміну зовнішньої та внутрішньої ситуацій на фінансовому ринку.

До основних принципів маркетингової стратегії в банку можна віднести:

- комплексність процесу роботи маркетингу (розуміння всіх співробітників стратегічних цілей та спрямування їх роботи на досягнення цих цілей);
- єдність перспективного та поточного планування маркетингу, що свідчить про наявність розроблених та діючих бізнес-технологій роботи з клієнтами;
- контроль прийнятих маркетингових рішень, проведення постійної оцінки якості обслуговування клієнтів.

Існує чотири основних маркетингових стратегій, вони наведені в табл.1.1

Таблиця 1.1 –Типи маркетингових стратегій

Стратегія	Характеристика
Стратегія росту	<ul style="list-style-type: none"> - продаж існуючих послуг на існуючих ринках; - продаж нових послуг на існуючих ринках; - продаж існуючих послуг на нових ринках; - продаж нових послуг на нових ринках;
Стратегія конкуренції	<ul style="list-style-type: none"> - лідерство по тарифам; - побудова клієнтських ланцюгів; - ексклюзивне обслуговування окремої групи клієнтів; - збільшення каналів розповсюдження інформації; - збільшення витрат; - використання прорахунків конкурентів.
Стратегія лідерства	<ul style="list-style-type: none"> - вирішення проблем внутрішньої комунікації; - покращення якості обслуговування клієнтів; - створення клієнтоорієнтованої організації.
Стратегія пріоритетів	<ul style="list-style-type: none"> - по послугам та збільшенню їх асортименту; - по ринковій ніші;

	<ul style="list-style-type: none"> - по існуючим клієнтам; - по зовнішньому комерційному середовищу;
--	--

Після визначення структури ринку банк вибирає для себе ту чи іншу стратегію маркетингу для досягнення поставлених цілей. Вибір стратегії визначається з одного боку зовнішніми умовами на ринку фінансових послуг та стратегією середньострокового розвитку банку з іншої.

На рис 1.1 представлені ключові фактори, визначаючі вибір маркетингової стратегії банку.



Рисунок 1.1 – Фактори, визначаючі вибір маркетингової стратегії банку

Ефективність реалізації маркетингової стратегії банку залежить від застосованих маркетингових інструментів. До основних маркетингових інструментів відносять концентрований, диференційний та масовий маркетинг.

Концентрований маркетинг (концентрація на одному сегменті ринку, коли банк концентрується на обслуговуванні одного сегменту споживачів). Дана концепція використовується для виходу на нові ринки.

Перевагою є деяка незалежність від дій конкурентів, концентрація на потребах клієнтів певного сегменту, ціленаправленість рекламних кампаній, стабільність прибутків банку. До недоліків відносять складний

процес боротьби з конкурентами за той самий сегмент ринку на перших стадіях, концентрацію на задоволенні потреб тільки певної групи споживачів, результатом чого є одноманітність банківських послуг, підтримку постійного контакту зі споживачами.

Дану концепцію використовують при обмежених ресурсах, для дрібних банків. Банк вкладає ресурси в ті послуги, де він має переваги у обслуговуванні перед іншими, великими банками, тим самим лідирує за рахунок унікальності та індивідуальності в задоволенні потреб клієнтів і економить свої ресурси. Стратегія концентрованого маркетингу дозволяє банку максимізувати прибуток на одиницю продукції, з невеликими ресурсами конкурувати з великими банками на спеціалізованих ринкових сегментах.

Іншою концепцією концентрованого маркетингу є задоволення однієї потреби всіх груп користувачів. Цю стратегію, як правило, обирають нещодавно створені банки, яким необхідно виділитись серед великої кількості інших банків. В основі вибору даної концепції лежить вибір такої спеціалізації, за якої послуга буде задовольняти широке коло споживачів і в той самий час надавати банку перевагу над існуючими конкурентами.

Перевагами даної стратегії є концентрація на детальному вивченні послуги, що може вдовольнити потреби широкого кола клієнтів і, як результат, банку буде надавати послуги набагато вищої якості, ніж його конкуренти. До недоліків, чи складнощів реалізації даної стратегії відносять складність пошуку та впровадження такого продукту, що міг би конкурувати серед послуг-аналогів конкурентів. А також постійне вдосконалення даного продукту та досить суттєві затрати на рекламні компанії, необхідні щоб охопити широке коло потенційних користувачів послуги.

Диференційований маркетинг полягає у спрямуванні зусиль банку на два чи більше сегментів ринку. Позитивними аспектами даної стратегії є

досить швидке пристосування до ринкової ситуації, можливість диференціації при визначенні цінової політики в залежності від платоспроможності клієнтів. Також до плюсів відносять диференціацію ризиків у різні сегменти ринку. Недоліком є значне збільшення затрат на рекламні кампанії, покращення якості продуктів та ін. зразу у двох чи більше сегментах.

Масовий маркетинг – це розробка однієї програми для всіх сегментів ринку, тобто надання послуг споживачам, що входять до споживацьких груп. Така стратегія застосовується коли ринок банківських послуг є однорідним. Перевагами даного методу є досить великий обсяг ринку та легке залучення нових клієнтів, невеликі затрати на рекламні кампанії. Недоліками масового маркетингу є високий рівень конкуренції, труднощі з завоюванням лідируючих позицій на ринку, розсіювання зусиль та ресурсів на багато продуктів.

Перш ніж вибрати маркетингову стратегію необхідним кроком є оцінка ефективності діяльності конкурентів на ринку. Для такої оцінки необхідно порівняти стратегії банків за певними елементами відповідно до критеріїв, що наведені у табл. 1.2

Таблиця 1.2 –Елементи маркетингової стратегії банку

Елемент маркетингової стратегії	Критерії ефективності маркетингової діяльності
Продукт	<ul style="list-style-type: none"> – бренд банку; – широта та глибина продуктового портфеля; – інтегральний показник рівня якості продукту; – якість сервісу;

	<ul style="list-style-type: none"> – швидкість зміни обсягів продажів.
Ціна	<ul style="list-style-type: none"> – рівень цін; – гнучкість цінової політики; – фактори ціноутворення нових продуктів.

Політика збуту	<ul style="list-style-type: none"> – цільові ринкові сегменти – обсяг реалізації; – кількість робітників маркетингового; – департаменту банку – рівень кваліфікації маркетингового департаменту банку; – ефективність роботи філіальної сітки; – використання інструментів прямого маркетингу.
Маркетингові комунікації	<ul style="list-style-type: none"> – рівень рекламної діяльності (бюджет, види реклами, види ЗМІ, характеристики рекламних кампаній); – рівень та методи стимулювання продажів(скидки, премії, купони, лотереї та конкурси; розмір бюджету стимулювання); – використання персональних продажів (їх обсяг у загальному обсязі реалізації, оплата праці та стимулювання робітників). Використання інструментів зв'язків з громадськістю .

Кінцева ціль маркетингової стратегії у банку – створити переваги над конкурентами на ринку банківських послуг.

Отже, для визначення маркетингової стратегії керівництву банку необхідно визначити цілі, які необхідно досягти для створення переваг перед конкурентами, розробити довгострокову та короткострокову політику досягнення таких цілей.

Продуктова політика займає ключове місце в комплексі маркетингу та включає в себе прийняття рішень по формуванню продуктового асортименту банку, модифікацію чи виключення продуктів, організацію системи додаткових послуг.

Банківський асортимент може буди наступних типів:

- базовий;
- поточний;
- розширений.

Аналіз продуктового ряду зручно виконувати за допомогою матриці Клієнти-Послуги. Дані матриці наглядно відображають попит на різні види послуг, дають змогу виявити асортимент банківських продуктів для цільової аудиторії, визначити потенціал ринку для кожного банківського продукту окремо. Даний аналіз є основою для розробки продуктової стратегії банку.

Продуктова стратегія банку включає в себе:

- оптимізацію асортименту продуктів та послуг, що надаються банком, з позицій стадій життєвого циклу продукту;
- розробку нових продуктів;
- аналіз та пропозиція комплексних рішень для клієнта, що складається з декількох банківських продуктів;
- виключення банківських продуктів, за неефективності банківських послуг та низького попиту на них.

При розробці продуктової стратегії необхідно враховувати наступні етапи життєвого циклу банківського продукту:

- етап виходу на ринок (на даному етапі особлива увага приділяється якості продукту, саме якість, а не кількість реклами визначає чи буде клієнт користуватися даним продуктом; також важливим чинником є чесність, висвітлення всієї інформації про умови банківських послуг);
- етап росту (на цьому етапі банк має модифікувати продукт для максимального задоволення потреб клієнтів);
- етап зрілості (політика підтримки якості та утримання постійних клієнтів);
- етап спаду (на цьому етапі необхідно оновити продукт чи виключити його з банківського асортименту).

Важливим елементом продуктової політики є розробка інноваційного продукту. Такі продукти мають задовольняти наступні вимоги:

- бути прибутковими;
- мати унікальні властивості;
- мати мінімальний кредитний ризик.

Процес створення такого продукту включає наступні стадії:

- пошук та оцінка ідей;
- визначення маркетингових можливостей;
- розробка банківського продукту;
- тестові продажі;
- аналіз та вдосконалення результатів виявлених на тестовому продажі;
- виведення продукту на ринок;

Варто зауважити, що попит на банківські продукти є досить чутливим до законодавчої політики і цей момент необхідно враховувати на початковому етапі розробки.

Комунікаційна політика банку є також ключовою частиною банківської стратегії. Клієнт перш за все лояльний до тих банків, яким він довіряє, в яких йому надали не тільки стандартні послуги, а й консультаційні

з питань, що його цікавлять. Комунікативна політика включає в себе систему засобів взаємодії банку з клієнтами.

Згідно законодавства банк має надавати інформацію про свою діяльність, а саме:

- фінансові показники;
- відомості про організаційну структуру;
- асортимент банківських послуг;
- ціни на банківські послуги;

Комунікативна політика банку здійснюється за допомогою наступних каналів:

- персонал банку;
- реклама;
- Засоби масової інформації;
- відгуки клієнтів.

До пункту комунікативної через персонал банку відносяться спілкування у відділенні банку, що пов'язане з консультацією та наданням послуг, телефонні комунікативні, комунікативні через чати, соціальні мережі, поштою та інші канали. Як правило, при таких видах комунікативної важливим фактором є компетенція працівника банку, його розуміння цілей та вміння доносити думку до клієнта.

Також банківський працівник має завоювати довіру клієнта. Через абстрактність банківських послуг та складності їх розуміння від банківських працівників вимагається висока професійна кваліфікація. Однією з ключових проблем особистих продажів банківських послуг є спосіб мислення працівників та їх мотивація.

Як уже зазначалося, реалізація маркетингової стратегії вимагає формування маркетингової філософії, тобто такого способу мислення, при якому всі банківські співробітники - від менеджерів до виконавців - у всіх своїх діях виходили б з потреб клієнтів. Формування маркетингової філософії

- тривалий процес, оскільки воно передбачає зміну культури і мислення, яке не можна здійснити насильно. У той же час цей процес можна полегшити шляхом створення для працівників відповідної мотивації.

Мотивація може формуватися як шляхом роз'яснення співробітникам суті проведеної банківським інститутом стратегічної лінії і їх місця в її реалізації, так і створенням атмосфери причетності до досягнення загальних цілей, а також відповідною організацією оплати праці. У систему оплати праці в зарубіжних банках, як правило, входять: відшкодування витрат, що виникають в ході надання послуг, зафіксована ставка з можливістю підвищення за допомогою просування по службі, премії, комісійні винагороди.

Однак використання комісійних винагород наштовхується і на певні проблеми: прагнення співробітників до отримання комісійних (наприклад, при консультуванні з приводу вкладення капіталу) може вступати в протиріччя з вимогами якісного обслуговування (об'єктивністю консультування). Особливо це відноситься до позаштатним співробітникам (самостійним посередникам), яких важко проконтролювати.

1.3 Сучасний підхід до формування банківської стратегії для кожного цільового ринку

На сьогодні в банківських установах підрозділи недостатньо синхронізовані у питаннях продуктової та комунікаційної політики. На етапі розробки маркетингових кампаній відбувається сегментування бази клієнтів за певними ознакам, що в подальшому допоможуть виокремити цільову аудиторію для банківських продуктів.

Сегментування ринку — розподіл потенційних споживачів на групи на основі відмінностей у їх потребах, характеристиках та поведінці. Застосування концепції ринкового сегментування дозволяє банку досягти максимальної результативності маркетингової діяльності шляхом використання своїх сильних сторін з урахуванням реальних умов на ринку.

Виокремлення сегментів на ринку ґрунтується на відборі важливих характеристик клієнта, наприклад:

- вікова група;
- рівень достатку;
- сімейне положення;
- наявність майна, тощо.

Результатом проведення маркетингових кампаній є збільшення потенційних клієнтів у відділеннях банків, де вони проходять процедуру скорингу.

Під кредитним скорингом розуміють визначення кредитоспроможності клієнта, шляхом нарахування йому балів за певними ознаками, та визначення нижньої границі для рішення по видачі кредиту чи продажу банківського продукту.

Скорингова модель, що не узгоджена та відокремлена від сегментації, проведеної маркетинговим підрозділом призводить до ситуацій, коли цільова аудиторія продукту згідно з маркетинговою стратегією не співпадає з результатами скорингу і потенційні клієнти отримують відмову. Це в свою чергу впливає на лояльність клієнтів, попит на продукти банку на всіх етапах життєвого циклу банківського продукту зокрема та на довіру до банку загалом. Узгодженість між цими двома моделями особливо важлива на початкових етапах, коли основним каналом залучення клієнтів є маркетингові та рекламні кампанії.

1.4 Скоринг

Кредитний скоринг (credit scoring, від англ. score – рейтинг) або аплікаційний скоринг (application scoring) – це методологія оцінювання кредитоспроможності потенційних позичальників у ризик-менеджменті [4], або методологія класифікації потенційних клієнтів (контрагентів) банку по ступеню (рівню) ризику [2], або набір моделей прийняття рішень та основоположних технік, що допомагають кредиторам в процесі вирішення питання надання споживчого кредиту [3]. Скоринг – методологія оцінювання кредитоспроможності або майбутньої поведінки на рівні клієнтів або договорів, як потенційних, так і існуючих [4], тому існує багато категорій скорингу: кредитний (аплікаційний) скоринг, поведінковий скоринг, скоринг виявлення та попередження шахрайства, колекторський скоринг, інші численні категорії скорингу [3, 8–11].

1.5 Скорингові моделі

Скорингова модель – це прогнозна модель визначення ризику дефолту позичальника, основана на історичних даних про клієнта та культуру виплат по конкретному кредиту.

У зв'язку зі збільшенням попиту на кредитні продукти для швидкого задоволення потреб клієнтів необхідно було зменшити час обробки заявки на кредит, не збільшуючи при цьому кредитні ризики. Для вирішення цього питання в 1956 році компанія Fair, Isaac & Co. (зараз відома як FICO) запропонувала першу скорингову систему.

Після обробки історичних даних можна було виокремити характеристики потенційних позичальників, що позитивно чи негативно впливали на результат взаємовідносин з банком. Далі, значенням характеристик присвоювали відповідні бали – що формували скорингову карту. При обробці анкети позичальника співробітник банку вносив дані в програму, які оброблялись за допомогою спеціального алгоритму. На виході позичальнику присвоювався скоринговий бал. Якщо бал відповідав внутрішнім політикам банку, заявка автоматично схвалювалась, що значно прискорювало та зменшувало роботу кредитного відділу банку.

З розвитком технологій, точність скорингових моделей значно покращилась. Наразі для побудови такої моделі аналітики використовують велику кількість статистичних, математичних методів та методів інтелектуального аналізу даних (ІАД).

Ці та інші фактори дозволили значно збільшити рівень автоматизації процесу прийняття рішень, що в свою чергу дозволило істотно зменшити час, необхідний для прийняття рішень.

Наразі кожен банк використовує певне програмне забезпечення або власні засоби автоматизації та підтримки процесу прийняття рішень на основі не тільки інформації, а й історичних знань, отриманих з внутрішніх або зовнішніх баз даних, що містять дані щодо власної історії кредитування, клієнтів, транзакцій, опису спеціальної інформації, інтерпретацій, внутрішньої термінології та ін.

1.6 Системи підтримки прийняття рішень

Система підтримки прийняття рішень (СППР) (англ. Decision Support System - DSS) - інтерактивна комп'ютерна автоматизована система

(програмний комплекс), що призначена для допомоги та підтримки різних видів діяльності людини при прийнятті рішень стосовно розв'язання слабо структурованих або неструктурованих проблем.

Застосування СППР забезпечує виконання ґрунтовного та об'єктивного аналізу предметної області при прийнятті рішень в складних умовах [4].

СППР застосовують у випадках коли особі, що приймає рішення для вирішення певного питання чи проблеми необхідне комплексне уявлення та аналіз великої кількості даних з предметної області.

Такі рішення можуть мати різну періодичність та різнитись за кількістю вхідних параметрів. Як правило, СППР розробляють як для державних установ, установ банківської системи, страхових компаній так і для малих суб'єктів підприємницької діяльності.

Системи підтримки прийняття рішень можуть бути як коробочне рішення, та не вимагати певних налаштувань від користувачів, так і повністю кастомізоване програмне забезпечення розроблене для конкретного підприємства чи установи з урахуванням специфіки його бізнес процесів та потреб.

Для прийняття правильного рішення ОПР має враховувати моменти суб'єктивної оцінки, прийняття однобічного рішення, відсутність врахування чи недооцінка ризику, неактуальністю рішення, неправильним тлумаченням статистичних даних та інші. СППР не використовується для повної автоматизації процесу, а лише виступає допоміжною системою для ОПР у момент прийняття рішення. Такі системи допомагають менеджерам вирішити певну задачу в умовах невизначеності та відсутності достатньої чи повної інформації про той чи інший процес.

Причиною виникнення і впровадження даних систем по всьому світу є зростаюча роль інформаційних технологій у всіх галузях діяльності людства, збільшенні кількості даних, глобалізацію. Сучасний рівень технологій

дозволяє досить швидко та якісно обробляти велику кількість даних та проводити складні обчислення.

Сучасна система підтримки прийняття рішень має бути легка у використанні, легко модифікуватись та змінюватись в залежності від бізнес процесів, допомагає, а не приймає рішення за ОПР. У багатьох країнах світу системи підтримки прийняття рішень розробляють та впроваджують корпорації, що спеціалізуються на розробці програмного забезпечення та консалтинговій діяльності.

СППР підтримує процес прогнозування в реальному часі на основі використання історичних даних, що накопичуються в базі даних. В результаті ОПР в реальному часі може спостерігати ситуацію і впливати на неї, а не лише аналізувати певні історичні факти. Наприклад, спеціаліст маркетингового відділу під час реалізації маркетингової кампанії відслідковує певні показники (опрацьовані та представлені в СППР) і вже на середні періоду кампанії бачить ефективність різних каналів маркетингової кампанії. Таким чином він може вирішити зупинити розповсюдження реклами через певний канал з метою зменшення витрат, та з огляду на його малоефективність.

На сьогодні актуальним завданням у розробці та впровадженні систем прийняття рішень є її створення на базі методів математичного моделювання та сучасних методів штучного інтелекту. Проте слід розуміти, що така система не повинна контролювати процеси без втручання та контролю людини, має підвищувати ефективність прийнятих рішень та бути простою у використання для користувачів.

1.7 Етапи проектування СППР

Проектування системи підтримки прийняття рішень в основному складається з таких етапів :

- декомпозиція процесу ухвалення рішення на прості операції та їх опис;
- опис функціональних вимог до системи;
- докладна специфікація функцій СППР.

На першому етапі проектування системи підтримки прийняття рішень необхідно визначитись з метою, який результат має бути після розробки даної СППР. Цей початковий крок є ключовим у процесі проектування системи підтримки прийняття рішень, адже без чітко визначеної мети використання СППР ніколи не вдасться створити необхідну для даного процесу систему. Кінцеві користувачі та всі особи, що будуть займатись даною розробкою мають чітко розуміти, яка система має бути в результаті впровадження.

Весь процес прийняття рішення поділяють на певні частини чи етапи. Далі в ході інтерв'ю з особами приймаючими рішення на кожному з етапів необхідно докладно описати процес прийняття рішення на момент до впровадження СППР, тобто «as is», то як це має відбуватися в майбутньому за допомогою системи - «to be». Варто намагатись зрозуміти процес прийняття рішень з декількох сторін: зі сторони особи що приймає рішення, та виходячи з зовнішнього розуміння проблеми, що стоїть перед ОПР.

Першим завданням є визначити ті задачі, які при передачі їх у виконання СППР не викривлять майбутнє рішення ОПР у гіршому напрямку. Ця задача є дуже складною та важливою у процесі побудови СППР, адже метою її впровадження є спрощення самого процесу прийняття рішень та збільшення ефективності прийняття рішень ОПР. Спочатку необхідно виконати декомпозицію мети, тобто виокремити певні цілі, що планується досягти після прийняття певного рішення. Прикладом таких цілей є підвищення доходу до планового рівня, досягнення певних характеристик системи, збільшення обсягу користування продуктом тощо. Далі необхідно

визначити задачі, які мають бути виконані для досягнення кожної з таких цілей.

Для докладного описання кроків з прийняття рішень використовують методику опитування чи інтерв'ю. На інтерв'ю мають бути присутні не тільки безпосередні особи, що приймають рішення, а й технічні спеціалісти, що забезпечують чи будуть забезпечувати даний процес. Після кожного такого інтерв'ю готується протокол з описанням процесу прийняття рішень. Протоколи обов'язково мають бути узгоджені всіма відповідальними особами, які приймають участь у процесі прийняття рішень. Покрокове описання ситуацій з прийняття рішень дозволяє знайти ті місця в процесі, які краще делегувати СППР. Така формальність забезпечує чітку постановку цілей СППР та мінімальну кількість розходжень планової архітектури та фактичної. Дані протоколи в кінці оформлюються в документ функціональних вимог до майбутньої системи підтримки прийняття рішень, що береться за основу для вибору технологій, які будуть використанні при розробці системи та написання специфікації. Також необхідно визначити інформаційне забезпечення процесу прийняття рішень та питання накопичення та використання даної інформації.

Наступним важливим кроком в процесі проектування СППР є аналіз і визначення обмежень, які можуть виникнути при прийнятті рішень. На цьому етапі слід визначити перелік труднощів ОПР в процесі прийняття рішень та можливість делегування цих функцій системі СППР, іншими словами розподіл міні задач по прийняттю рішень між ОПР та СППР.

Важливо чітко розуміти, які задачі можна делегувати СППР і з якими вона впорається краще за ОПР, а які все ж таки необхідно вирішувати безпосередньо людині. Прикладом таких задач є наступні:

- Складність у реалізації прогнозування
- Складність із порівнянням атрибутів і цілей

- Складність у опрацюванні великих масивів даних та розрахунку показників, використовуючи алгоритми
- Складність з візуалізацією інформації та формуванні висновків
- Складність з пошуком потрібної інформації

Всі перелічені складності можна спростити шляхом використання електронно-обчислювальної техніки.

На останньому етапі встановлюється відповідність між бажаними функціями системи підтримки прийняття рішень та методами, що підтримують реалізацію цих функцій і співпадають з процесом прийняття рішень в конкретній ситуації.

Важливо розуміти, що вибір методів виходить з поняття ефективності та доцільності застосування саме цього методу для виконання певної задачі. Після завершення цього етапу формується єдина функціональна архітектура системи підтримки прийняття рішень.

Наступним етапом є докладне проектування інтерфейсу майбутньої системи підтримки прийняття рішень, а також її структури керування. Основною вимогою є легкість використання, інтуїтивна зрозумілість інтерфейсу, наявність підказок та статей баз знань.

На цьому моменті закінчується опис вимог до системи, та починається створення програмної системи і її тестування.

Інструментальний підхід до проектування СППР- генераторів.

СППР-генератор- це сукупність взаємопов'язаного програмного забезпечення, використання якого спрощує побудову системи підтримки прийняття рішень.

Наприклад, комплекс програмного забезпечення, що поєднує в собі програмами для пошуку, обробки та збереження інформації, аналітики та формування звітів, засобів для виконання фінансових та статистичних аналізів. Перевагою даної системи є можливість її створення та модифікації не програмістами.

Висновки до розділу

В умовах підвищеної конкуренції та однорідності продуктів банківській сфері банківський маркетинг стає однією зі сфер розвитку банківських установ. Основним фактором, на який звертають увагу клієнти є довіра до банківської установи. Банківський маркетинг, що має на меті збільшення довіри до банку та лояльності клієнтів – ось одне з головних завдань сучасних банків.

Лояльність до банку досягається шляхом винесення потреб клієнтів на перший план та надання їм спеціального обслуговування. Застосування інформаційних технологій може зробити цей спосіб функціонування набагато продуктивнішим, а банк конкурентоспроможним на ринку фінансових послуг. Наразі у кожному банку існує база даних, яка містить важливу інформацію про клієнтів. Правильна обробка таких даних, побудова моделі, аналіз та формування відповідних висновків може надати банку конкурентну перевагу, а споживачам- якісне обслуговування. Метою даної роботи є розробка системи підтримки прийняття рішень для впровадження маркетингової стратегії у банку за допомогою оцінювання та прогнозування кредитоспроможності населення у режимі реального часу використовуючи скоринговий метод. Цілями даної системи підтримки прийняття рішень є :підвищення лояльності та довіри клієнтів банку, зниження витрат на маркетингові кампанії, зменшення кількості відмов у наданні кредитів, зменшення кількості заборгованості, збільшення операційної ефективності.

РОЗДІЛ 2 ВИБІР І ОПИС МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЛЯ СКОРИНГУ

2.1 Різновиди машинного навчання

У відповідь на запитання "Який алгоритм машинного навчання обрати?" як правило відповідають: "Це залежить від ряду обставин". Вибір залежить від розміру, якісних характеристик та природи даних. Навіть найдосвідченіші фахівці з даними не зможуть визначити найкращий алгоритм, не спробувавши його.

Для спрощення вибору алгоритму компанія Microsoft запропонувала наступний алгоритм, що зображений на рис.2.1

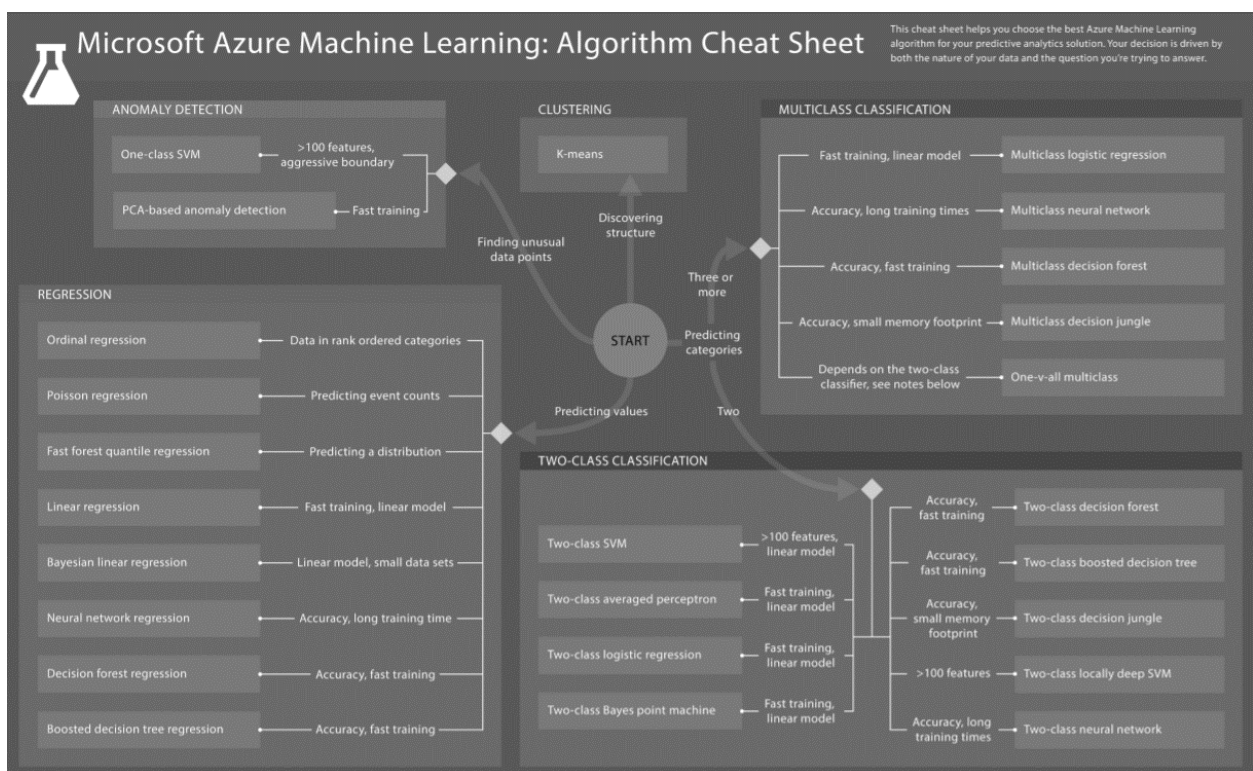


Рисунок 2.1 – Схема вибору алгоритму машинного навчання

Позначення шляху і алгоритму на схемі слід читати як "для <позначення шляху> використовуйте <алгоритм>". Наприклад, "для швидкості використовуйте двокласну логістичну регресію". Іноді може

використовуватися більше однієї гілки алгоритму. Іноді жодна з гілок алгоритму не підходить ідеально. Ці рекомендації наближені.

Існують наступні види алгоритмів машинного навчання:

- неконтрольовані;
- контрольовані;
- алгоритм з підкріпленням.

Контрольовані алгоритми навчання виконують прогнозування на основі наборів даних з готовими відповідями. Наприклад, вартість акцій за минулий період дозволяє припустити вартість акцій в майбутньому. Кожен приклад, який використовується, позначається цільовим значенням - в конкретній ситуації це вартість акції.

Контрольований алгоритм виконує пошук шаблонів для цих значень. Він використовує деякі відомості - день тижня, пора року, фінансова статистика організації, тип галузей, історичні події - алгоритм формує шаблони різних типів. Коли алгоритм виявив найкращий шаблон, він може на основі цього шаблону передбачити перевірочні дані - завтрашні ціни.

Контрольоване навчання - це досить цікавий вид машинного навчання.

Виділяють кілька типів контрольованих алгоритмів навчання:

- за допомогою класифікації;
- за допомогою алгоритму регресії;
- спрямовані на пошук аномальних ситуацій.

Якщо дані використовують для прогнозу певної ознаки, контрольоване навчання також має назву класифікація. Така ситуація відбувається, коли малюнок визначається як відображення "кішки" чи "собаки". При наявності тільки двох варіантів така класифікація називається двокласною або біноміальною. При наявності декількох категорій, класифікація називається багатокласовою.

При прогнозуванні показника, прикладом може бути вартості акцій, контрольоване навчання має назву регресії.

Іноді завдання полягає в ідентифікації точок даних, які просто є незвичайними. Прикладом може бути ідентифікація ситуацій шахрайства з кредитною карткою підозрілими є будь-які незвичайні операції оплати. Кількість імовірних випадків велика, а відомих прикладів дуже мало для того щоб навчитися визначати шахрайські дії дуже важко. Підхід, який використовується при виявленні аномалій, полягає в тому, щоб просто вивчити, як виглядають нормальні дії (за допомогою журналу нормальних транзакцій), і визначати всі дії, які суттєво відрізняються від нормальних.

При неконтрольованому навчанні точкам даних не присвоюються мітки. Замість цього мета алгоритму неконтрольованого навчання - певне упорядкування даних або опис їх структури. Це може означати угруповання даних в кластери або пошук різних способів аналізу складних даних для спрощення чи поліпшення їх організації.

У навчанні з підкріпленням алгоритм обирає дію, що є відповіддю на кожну точку інформації. Алгоритм також незабаром отримує сигнал, що сповіщає про успіх, який дає зрозуміти, як вдало було прийнято рішення.

На основі цього алгоритм змінює кроки навчання для досягнення кращого результату. Навчання з підкріпленням широко використовується в робототехніці, де сукупність показників датчиків в один момент часу являє собою точку даних і необхідно обрати наступну дію робота.

2.2 Параметри, що необхідні для визначення алгоритму

Перед початком вибору алгоритму слід визначити ті параметри, що найбільш прийнятні для майбутньої моделі. До таких параметрів відносять:

- точність;
- час навчання;
- лінійність;
- кількість параметрів алгоритму;
- кількість функцій;
- особливі випадки, тощо.

Отримання найбільш точної відповіді необхідно не завжди. Іноді досить наближеного відповіді в залежності від того, для чого він використовується. В цьому випадку можна значно скоротити час обробки, дотримуючись більш наближених методів. Ще однією перевагою більш наближених методів є те, що вони природним чином прагнуть уникнути надмірно високої точності.

Кількість хвилин або годин, необхідних для навчання моделі, сильно відрізняється для різних алгоритмів. Час навчання часто тісно пов'язане з точністю - одне зазвичай супроводжує іншому. Крім того, деякі алгоритми більш чутливі до кількості точок даних, ніж інші. Коли час обмежений, це може вплинути на вибір алгоритму, особливо з великим набором даних.

Лінійність використовується в багатьох алгоритмах машинного навчання. Алгоритми лінійної класифікації припускають, що класи можуть бути розділені прямою лінією (або її аналогом для більшого числа вимірів). До них відносяться логістична регресія і метод опорних векторів. Алгоритми лінійної регресії припускають, що тренди даних слідуєть прямій лінії. Ці припущення допустимі для ряду завдань, але для інших завдань вони призводять до зниження точності.

Параметри є тими регуляторами, які фахівець з даними підбирає при налаштуванні алгоритму. Це числа, які впливають на поведінку алгоритму, наприклад чутливість до помилок або кількість ітерацій, або на варіанти поведінки алгоритму. Час навчання і точність алгоритму іноді можуть бути дуже чутливими до точності завдання параметрів. Як правило, алгоритми з великим числом параметрів вимагають великої кількості проб і помилок, щоб визначити вдале поєднання параметрів. Плюсом є те, що наявність великої кількості параметрів зазвичай означає, що алгоритм має велику гнучкість. Це часто дозволяє досягати дуже хорошою точності. За умови, що ви можете знайти правильну комбінацію параметрів.

Для деяких типів даних кількість функцій може бути дуже велика у порівнянні з кількістю точок даних. Це часто відбувається з генетичними або текстовими даними. Велика кількість функцій для деяких алгоритмів навчання може привести до того, що вони загрузнуть і час навчання стане неприпустимо великим. Для цього варіанту особливо добре підходить метод опорних векторів.

Деякі алгоритми навчання роблять певні припущення про структуру даних або бажаних результатів. Якщо ви зможете знайти той алгоритм, який відповідає вашим потребам, з ним ви зможете отримати більш точні результати, більш точні прогнози і скоротити час навчання.

2.3 Вибір алгоритмів машинного навчання

Для задачі визначення рівня кредитоспроможності підходять алгоритми двохкласової класифікації, що наведені в табл.2.1

Таблиця 2.1 - Алгоритми двохкласової класифікації

Алгоритм	Точність	Час навчання	Лінійність	Кількість параметрів	Примітки
Логістична регресія		●	●	5	
Ліс рішень	●	○		6	
Джунглі рішень	●	○		6	Малий об'єм пам'яті
Дерево рішень	●	○		6	Великий об'єм пам'яті
Нейронна мережа	●			9	
Персептрон	○	○	●	4	
Метод опорних векторів		○	●	8	Для великих наборів функцій
Точкова машина Байєса		○	●	3	

Властивості алгоритму:

- - показує високу точність, швидкий час навчання і використовує лінійність
- - показує хорошу точність і помірне час навчання

Для скорингової моделі, що буде використовуватись у СППР для впровадження маркетингових стратегій у банку найбільш важливими характеристиками є точність та час навчання алгоритму.

Точність моделі відіграє велику роль у об'ємі вибірки цільової аудиторії для майбутньої маркетингової кампанії, а отже на пряму впливає на обсяг витрат на маркетингову діяльність банку та майбутній дохід від впровадження маркетингових кампаній.

Важливість швидкості навчання алгоритму визначається направленістю роботи системи підтримки прийняття рішень, яка має найшвидше реагувати на запити користувачів, в нашому випадку працівників маркетингового відділу.

Для вирішення поставлених задач найбільше підходять алгоритми логістичної регресії та дерева рішень.

2.4 Лінійні регресійні моделі оцінювання кредитоспроможності клієнтів за скоринговим методом

Для кредитного скорингу використовуються декілька класів моделей, що ґрунтуються на принципово різних математичних методах та підходах. Розглянемо найбільш поширені класи більш детально.

Логістична регресія - це алгоритм класифікації. Він використовується для прогнозування бінарного результату (1/0, Так / Ні, True / False) з урахуванням багатьох незалежних змінних. Для прогнозування результату вводиться змінна y , що набуває значення 0 або 1 та безліч незалежних змінних $(x_1, x_2 \dots)$. Незалежні змінні називають предикторами чи регресорами. Також вводиться фіктивний регресор $x_0=1$.

Далі робиться припущення, що $y=1$. Вірогідність даного результату розраховуються за формулою (2.1):

$$P\{y = 1|x\} = f(z), \quad (2.1)$$

$$\text{де } z = \theta^T x = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n, x$$

θ – вектори незалежних змінних $1, x_1, \dots, x_n$

$f(z)$ – логістична функція

Логістична функція представлена формулою (2.2)

$$f(z) = \frac{1}{1+e^{-z}} \quad (2.2)$$

Логістична функція представлена сигмоїдом на рис.2.2

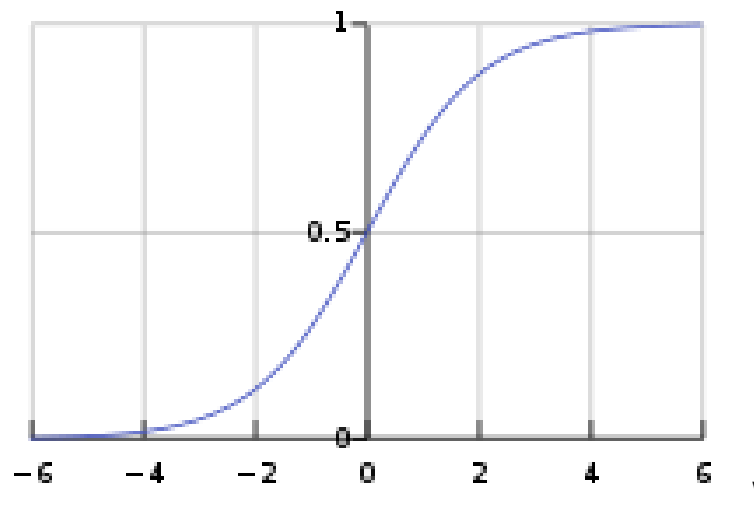


Рисунок 2.2 – Логістична функція

Так як y може бути або 0 або 1, ймовірність $y=0$ вираховується за формулою (2.3):

$$P\{y = 0|x\} = 1 - f(z) = 1 - f(\theta^T x) \quad (2.3)$$

де $z = \theta^T x = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n$, x

θ – вектори незалежних змінних $1, x_1, \dots, x_n$

$f(z)$ – логістична функція

Отже, функція розподілу y при заданому x можна записати в наступному вигляді формули (2.4):

$$P\{y|x\} = f(\theta^T x)^T (1 - f(\theta^T x))^{1-y}, y \in \{0,1\} \quad (2.4)$$

де $z = \theta^T x = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n, x$

θ – вектори незалежних змінних $1, x_1, \dots, x_n$

$f(z)$ – логістична функція

Для підбору параметрів $\theta_0, \dots, \theta_n$ необхідно створити навчальну вибірку, що складається з наборів незалежних змінних і відповідних значень залежної змінної y .

Для підбору параметрів використовують наступні методи:

- метод максимальної правдоподібності (МПП) (представлений формулою 2.5);
- метод Ньютона.

Відповідно до методу максимальної правдоподібності, підбираються такі параметри θ , які максимізують значення функції правдоподібності на навчальній вибірці.

$$\hat{\theta} = \operatorname{argmax}_{\theta} \prod_{i=1}^m P\{y = y^i | x = x^i\} \quad (2.5)$$

Де $\hat{\theta}$ – невідомий параметр,

$\prod_{i=1}^m P\{y = y^i | x = x^i\}$ – функція правдоподібності

Метод Ньютона (метод дотичних) – ітераційний чисельний метод знаходження кореня заданої функції.

Послідовні наближення методу Ньютона обчислюються за формулою (2.6):

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad x = 0, 1, 2 \dots \quad (2.6)$$

Де $x(0)$ – деяке початкове наближення до кореня.

Також може використовуватись метод стохастичного градієнтного спуску.

Припущення, зроблені логістичною регресією щодо розподілу та співвідношень у даних, майже такі ж, як припущення, зроблені в лінійній регресії. Багато досліджень було зроблено для визначення цих припущень та використовувалась точна ймовірнісна і статистична мова. Для правильного прогнозування перед використанням логістичної регресії необхідно підготувати дані у відповідності з наступними вимогами:

- видалити шум (викиди) у вхідній змінній;
- виконати перетворення даних вхідних змінних;
- видалити пов'язані вхідні дані;
- видалити пусті та нульові значення.

Необхідно обов'язково видалити шум: логістична регресія не передбачає жодних помилок у вихідній змінній (y), необхідно розглянути можливість вилучення викидів і, можливо, некласифікованих екземплярів із даних навчання.

Перетворення даних вхідних змінних: логістична регресія - це лінійний алгоритм (з нелінійним перетворенням на виході). Вона передбачає лінійне відношення між вхідними змінними і вихідним. Перетворення даних вхідних змінних може призвести до збільшення точності моделі

Видалення пов'язаних вхідних даних: як і лінійна регресія, модель логістичної регресії може перенавчитись, якщо у вибірці наявні висококорельовані вхідні дані. Необхідно розрахувати кореляцію між усіма парами вхідних параметрів та видалити всі висококорельовані.

Видалення пустих та нульових значень підвищить ймовірність збіжності.

Щоб оцінити ефективність логістичної регресійної моделі, ми повинні розглянути кілька показників, а саме:

- Akaike Information Criteria;
- залишки регресії (Residual Deviance);
- змішана матриця (Confusion Matrix).

AIC (Akaike Information Criteria) - аналогічна метрика коригованої R^2 в логістичній регресії є AIC. AIC - міра відносної якості економетричних (статистичних) моделей, що враховує ступінь «підгонки» моделі під дані з коригуванням (штрафом) на використану кількість оцінюваних параметрів. Тобто критерії засновані на якомусь компромісі між точністю і складністю моделі. Перевага завжди надається моделі з мінімальним значенням AIC.

Залишки регресії (Residual Deviance) - це різниця між фактичними значеннями і значеннями, передбаченими регресійною моделлю. Чим краще регресійна модель узгоджується з даними, тим менше величина залишків. Вираховуються залишки відповідно формулі (2.7):

$$e_i = (y_i - \hat{y}_i) \quad (2.7)$$

де y_i – фактичне значення,

\hat{y}_i – відповідне прогнозоване значення.

Змішана матриця - це табличне представлення фактичних чи прогнозованих значень, що допомагає нам знайти точність моделі та уникнути перенавчання (рис.2.3)

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Рисунок 2.3 – Змішана матриця

Точність моделі розраховуються за допомогою формули (2.8) :

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.8)$$

де TP – true positive – правильне прогнозування 1,

TN –true negative - правильне прогнозування 0

FN – false negative - неправильне прогнозування 0

FP – false positive - неправильне прогнозування 1

Коефіцієнт неправильної класифікації (error rate) розраховуються за наступною формулою (2.9):

$$ER = \frac{FP+FN}{TP+TN+FP+FN} \quad (2.9)$$

де TP – true positive – правильне прогнозування 1,

TN –true negative - правильне прогнозування 0,

FN – false negative - неправильне прогнозування 0,

FP – false positive - неправильне прогнозування 1.

Також можна розрахувати наступні показники:

- True Positive Rate (TPR) (Sensitivity; Recall);
- True Negative Rate (TNR) Specificity;
- False Positive Rate (FPR) ;
- False Negative Rate (FNR);
- Precision (Positive predictive value)(PREC);
- Коефіцієнт кореляції - Matthews correlation coefficient (MCC).

True Positive Rate (TPR) розраховується за формулою (2.10):

$$TPR = \frac{TP}{TP+FN} \quad (2.10)$$

де TP – true positive – правильне прогнозування 1,

FN – false negative - неправильне прогнозування 0

True Negative Rate (TNR) розраховується за формулою (2.11):

$$TNR = \frac{TN}{TN+FP} \quad (2.11)$$

де TN –true negative - правильне прогнозування 0,

FP – false positive - неправильне прогнозування 1

False Positive Rate (FPR) розраховується за формулою (2.12):

$$FPR = \frac{FP}{TN+FP} \quad (2.12)$$

де TN –true negative - правильне прогнозування 0,

FP – false positive - неправильне прогнозування 1.

False Negative Rate (FNR) розраховується за формулою (2.13):

$$FNR = \frac{FN}{TP+FN} \quad (2.13)$$

де TP – true positive – правильне прогнозування 1,

FN – false negative - неправильне прогнозування 0.

Precision розраховується за формулою (2.14):

$$PREC = \frac{TP}{TP+FP} \quad (2.14)$$

де TP – true positive – правильне прогнозування 1,

FP – false positive - неправильне прогнозування 1.

Коефіцієнт кореляції розраховується за формулою (2.15):

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (2.15)$$

де TP – true positive – правильне прогнозування 1,

TN –true negative - правильне прогнозування 0,

FN – false negative - неправильне прогнозування 0,

FP – false positive - неправильне прогнозування 1.

Показник F-score середнє гармонічне між показниками Precision and Recall. Проказник F1 досягає свого найкращого результату в значенні 1 та найгіршого в значенні 0, та розраховується за формулою (2.16):

$$F_1 = \frac{2*PREC*REC}{PREC+REC} \quad (2.16)$$

ROC-крива - графік, що дозволяє оцінити якість бінарної класифікації, відображає співвідношення між об'єктами об'єктів з загальною кількістю носіїв ознак TPR і долі об'єктів від загальної кількості об'єктів, ознак FPR. Аналіз класифікації з застосуванням ROC-кривих називається ROC-аналізом. Приклад ROC-кривої наведений на рис.2.4.

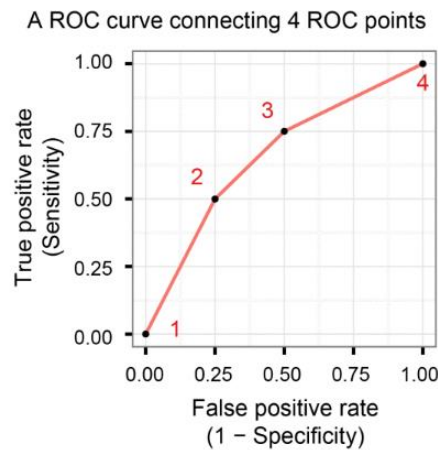


Рисунок 2.4. - ROC-крива

ROC-криві використовують для візуального порівняння точності моделей (рис.2.5).

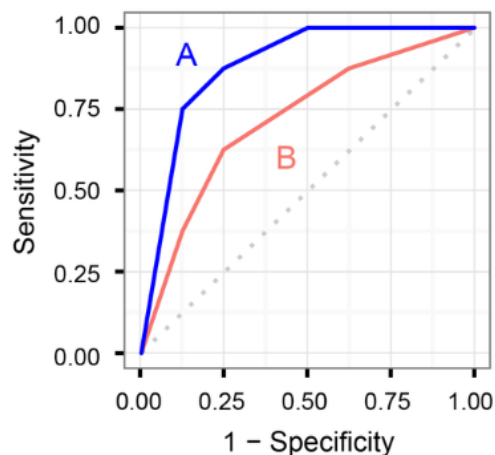


Рисунок 2.5. – Порівняння моделей за допомогою ROC-кривих

Коефіцієнт інтерпретації ROC дає показник AUC (англ. Area under the ROC curve, площа під ROC-кривою) - площа, обмежена ROC-кривою та

віссю долі FP. Чем більший показник AUC, тим точніша модель. При цьому значення 0,5 демонструє непридатність вибраного методу класифікації (відповідає випадковому вгадуванню). Значення менше 0,5 говорить, що класифікатор діє з точністю до навпаки: якщо позитивні назвати негативними і навпаки, класифікатор буде працювати краще.

2.5 Алгоритм дерев рішень оцінювання кредитоспроможності клієнтів

Одним з основних альтернативних методів класифікації у відношенні до регресійного методу моделювання є рекурсивно-partiційні алгоритми (recursive partitioning algorithms) у вигляді дерев класифікації (classification trees) – дерев рішень (decision trees) [5, 8, 9, 14].

Дерево рішень - є способом опису правил в ієрархічній, послідовній структурі, де для кожного об'єкта існує один вузол, що надає рішення.

Під правилом мається на увазі логічний вираз, представлений виразом "якщо ... ,то ...".

Створення дерев рішень. Нехай дано певну навчальну множину T , що містить об'єкти, кожен з об'єктів характеризується m атрибутами, а один з атрибутів визначає приналежність даного об'єкта до відповідного класу.

Нехай $\{C_1, C_2, \dots, C_k\}$ - класи, тоді можуть існувати три ситуації:

- множина T містить один або більше прикладів, які є представниками одного класу C_k . В такій ситуації дерево рішень для множини T - це лист, який визначає клас C_k ;
- множина T не містить ні одного прикладу. Тоді це також лист, і клас, пов'язаний з листком, обирається з іншої множини відмінної від T ;

- множина T має приклади, що відносяться до різних класів. В такому випадку необхідно розділити множину T на певні підмножини. Для цього обирається одна з ознак, що має два і більше відмінних один від одного зі значень $O_1, O_2, \dots O_n$. Множина T розбивається на підмножини $T_1, T_2, \dots T_n$, де кожна підмножина T_i має всі приклади, що мають значення O_i для вибраної ознаки. Це процедура триває рекурсивно поки кінцева множина не буде складатися із прикладів, що відповідає одному і тому ж класу.

Дана процедура покладена в основу основних алгоритмів створення дерев рішень, даний метод також відомий як метод розподілу та захвату (divide and conquer). При використанні цієї методики, створення дерев рішень буде відбуватись від верху до низу.

Через те що усі об'єкти були раніше визначені до існуючих класів, цей процес створення дерев рішень має назву навчання з вчителем (supervised learning). Процес навчання ще можна назвати індуктивним навчанням чи індукцією дерев (tree induction).

На сьогодні існує досить велика кількість алгоритмів, які будують дерева рішень.

- Алгоритм ID3, де вибір атрибута відбувається на підставі приросту інформації (англ. Gain), або на підставі критерію Джині.
- Алгоритм C4.5 (поліпшена версія ID3), де вибір атрибута відбувається на підставі нормалізованого приросту інформації (англ. Gain Ratio).
- Алгоритм CART і його модифікації - IndCART, DB-CART.
- Автоматичний детектор взаємодії Хі-квадрат (CHAID). Виконує багаторівневе поділ при розрахунку класифікації дерев;
- MARS: розширює дерева рішень для поліпшення обробки цифрових даних.

Багато алгоритмів є "жадібними алгоритмами". За умови що один раз був вибраний атрибут, і по цьому атрибуту було зроблено поділ на підмножини, то алгоритм не може повернутися назад і обрати інший атрибут, який допоміг би краще розподілити множину.

Розглянемо алгоритм C4.5, що є вдосконаленою версією алгоритму ID3 (Iterative Dichotomizer), використовує теоретико-інформаційні підходи.

Для вибору оптимального атрибута, запропоновано критерій за формулою (2.17):

$$\text{Gain}(X) = \text{Info}(T) - \text{Info}_x(T) \quad (2.17)$$

де $\text{Info}(T)$ – ентропія множини T .

Ентропія множини T представлена формулою (2.18):

$$\text{Info}_x(T) = \sum_{i=1}^n \frac{|T_i|}{|T|} * \text{Info}(T_i) \quad (2.18)$$

Множини T_1, T_2, \dots, T_n отримані при поділі вихідної множини T з перевірки X .

Обирається атрибут, що дає максимальне значення за критерієм.

Вперше ці дії були запропоновані Р. Куінленом в розробленому ним алгоритмі ID3. Крім вищописаного алгоритму C4.5, існує клас алгоритмів, які використовують цей критерій вибору атрибута.

Алгоритм CART використовує так званий індекс Gini, представлений формулою (2.19), який оцінює "відстань" між розподілами класів.

$$Gini(c) = 1 - \sum_j p_j^2 \quad (2.19)$$

де c - поточний вузол,

а p_j^2 - імовірність класу j в вузлі c .

До переваг алгоритму можна віднести:

- швидкий процес навчання;
- генерація правил в областях, де експерту важко формалізувати свої знання;
- витяг правил на природній мові;
- інтуїтивно зрозуміла класифікаційна модель;
- висока точність прогнозу, порівнянна з іншими методами.

До недоліків моделі можуть відносити занадто складні конструкції, які недостатньо повно представляють дані. Дана проблема називається перенавчанням. Для того, щоб уникнути цієї проблеми, необхідно використовувати Метод «регулювання глибини дерева».

Висновки до розділу

В даному розділі розглянуто основні підходи до побудови математичних моделей, на яких базуватиметься розробка скорингової моделі. До основних різновидів машинного навчання відносяться неконтрольовані алгоритми, контрольовані алгоритми та алгоритми з підкріпленням. Були визначені параметри, які необхідно визначити перед вибором алгоритму машинного навчання, а саме: точність, час навчання, кількість параметрів та кількість функцій. Розглянуто можливість застосування алгоритмів логістичної регресії, дерев, лісу та джунглів рішень, перептрону, нейронних мереж, алгоритму опорних векторів та точкової машини Байєса.

В якості основної моделі було запропоновано використання регресійної моделей, а саме, логістичної регресії та алгоритм дерев рішень. Ці моделі є найбільш розповсюдженими та найбільш простими для обчислень. Моделі на основі регресії не потребують значних обчислювальних ресурсів, а також, мають найбільш детальну методологію впровадження.

РОЗДІЛ 3 РОЗРОБКА СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ НА ОСНОВІ ПОБУДОВИ СКОРИНГОВИХ МОДЕЛЕЙ

3.1. Формулювання вимог до СППР

Система підтримки прийняття рішень для впровадження маркетингових систем у банку має допомогти спеціалісту маркетингового відділу банку підтвердити його гіпотезу стосовно сегменту та нового продукту, з точки зору кредитоспроможності майбутніх покупців продукту.

Дана система підтримки прийняття рішень в цілому має задовольняти наступні умови:

- інтерактивність;
- інтегрованість;
- потужність;
- доступність;
- гнучкість;
- надійність;
- керованість.

Основною задачею системи є узгодження маркетингової політики з політикою видачі кредиту банку. Головні вимоги з точки зору кінцевих користувачів – простота у використанні та інтуїтивний (зрозумілий) інтерфейс. Такі вимоги виходять на перший план за рахунок орієнтації системи підтримки прийняття рішень на спеціалістів маркетингового департаменту банку як кінцевих користувачів. Слід розуміти, що дані користувачі можуть не мати базових знань програмування чи мови запитів. Для них система виглядатиме як форма з полями для вводу та області відповідей.

Також користувач повинен мати можливість створити сегмент чи маркетинговий список після проведення скорингу в системі для подальшого його використання в маркетинговій кампанії. Система має підказувати користувачу, який відсоток дефолтів був у сегменті, який він обрав та рекомендувати чи не рекомендувати використовувати його в майбутній маркетинговій кампанії.

Цілями даної системи підтримки прийняття рішень є :

- підвищення лояльності та довіри клієнтів банку;
- зниження витрат на маркетингові кампанії;
- зменшення кількості відмов у наданні кредитів;
- зменшення кількості заборгованості ;
- збільшення операційної ефективності .

3.2 Розробка архітектури та функціональної схеми програми

У ході даної роботи була розроблена наступна архітектура системи (рис.3.1) підтримки прийняття рішень для впровадження маркетингової стратегії у банківській установі.

На першому етапі користувач системи (працівник маркетингового відділу) вносить параметри продукту для якого планується проведення маркетингової кампанії та параметри сегменту для даного продукту. Параметри сегменту залежать від даних конкретного банку. До стандартних параметрів відносять:

- купівельна поведінка;
- територія;
- соціальний статус;
- стиль життя;
- продукт (наприклад, вид кредиту);

- вік;
- стать;
- посада
- освіта;
- наявність авто, нерухомості;
- дохід.

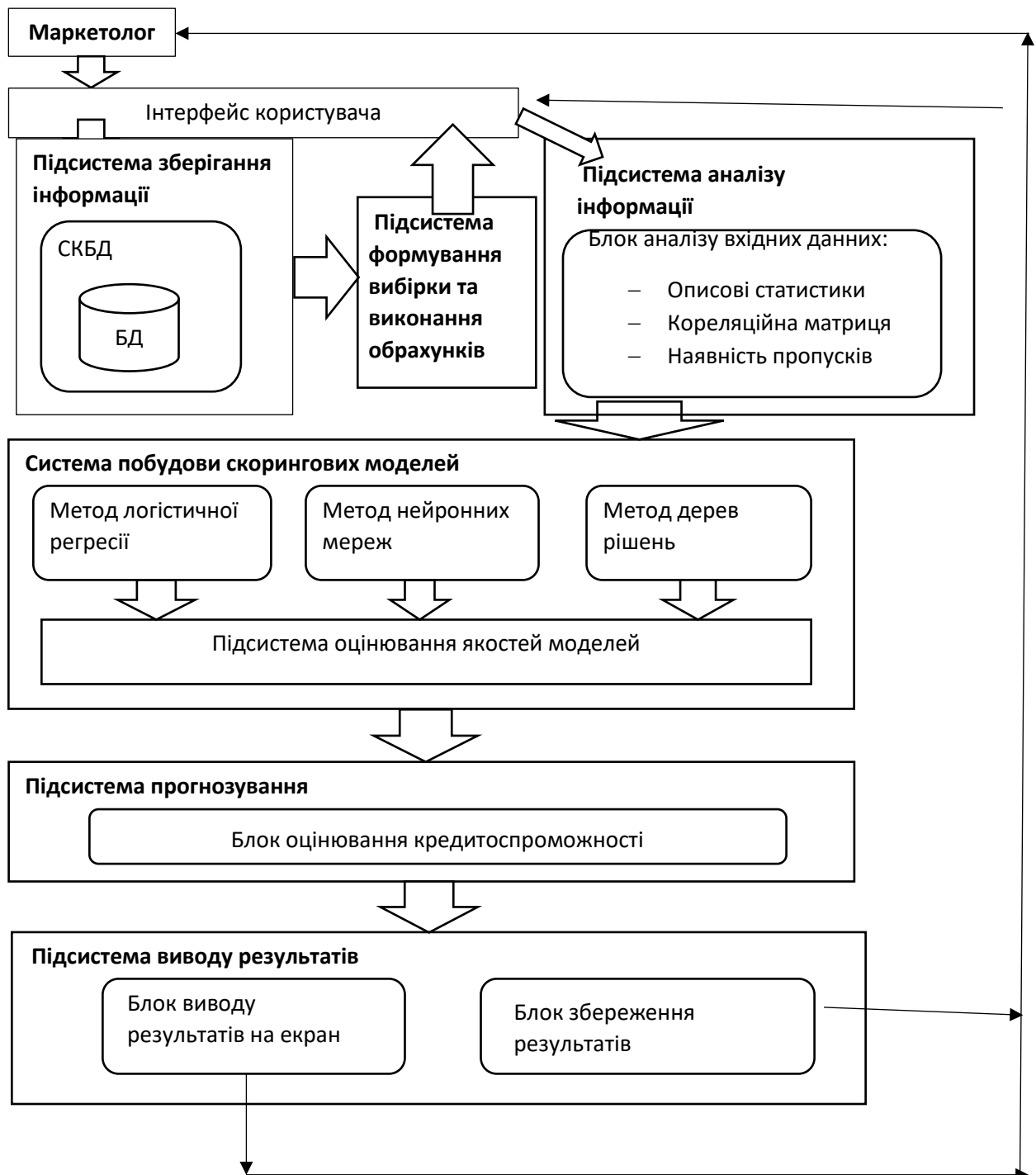


Рисунок 3.1 – Архітектура розробленого програмного продукту

До параметрів продукту відносять:

- максимальна сума тіла кредиту;
- мінімальна сума тіла кредиту;
- термін кредиту;
- відсоток;
- комісії;
- можливість застави та поручительства, тощо.

Дані параметри можуть змінюватись в залежності від продукту і як результат від сегменту маркетингової кампанії. В результаті обробки великої кількості даних по кредитній історії клієнтів банку (та можливо кредитних історій, що були отримані від Бюро Кредитних Історій) система показує результат, а саме :

- відсоток дефолтів у даному сегменті;
- рекомендація до використання даного сегменту у подальшій маркетинговій кампанії.

Після прийняття рішення ОПР, про застосування даного сегменту, система опрацьовує дані вибраного сегменту. Результатом скорингу є скорингові карти, які застосовуються безпосередньо у відділеннях менеджерами для визначення ризику та прийняття рішення про надання кредиту. Після цього формується вибірка з існуючих та потенціальних клієнтів та передається у CRM-систему для подальшого використання, а саме виконання маркетингових активностей (наприклад, розсилки sms).

3.3 Вибір інструментальної платформи для виконання обчислювальних експериментів

Для проведення експериментів та вибору найкращої моделі, що буде використана у системі підтримки прийняття рішень було обрано студію машинного навчання Azure розроблену компанією Microsoft.

Студія машинного навчання Microsoft Azure - це інструмент для обробки та аналізу даних, що підтримує функцію інтерактивної взаємодії з об'єктами для створення, тестування та впровадження рішень для прогностного аналізу даних. Microsoft Azure представляє інтерактивний візуальний робочий простір, що спрощує створення, тестування та виконання ітерацій моделі прогностної аналітики. Пересуваючи набори даних та модулі аналізу на інтерактивну робочу область і зв'язуючи їх разом створюється експеримент (модель), який потім виконується в студії. Даний експеримент можна редагувати та виконувати знову і знову. Після формування навчаючого експерименту його можна перетворити в прогностний, а потім опублікувати як веб-службу, для того щоб модель стала доступна для інших користувачів.

Перевагою даного рішення є можливість публікації веб-служб, які потім можна використовувати в користувацьких застосунках та засобах бізнес-аналітики. Процес створення навчального експерименту наведений на рис. 3.2.



Рисунок 3.2 - Процес створення навчального експерименту

На етапі імпорту даних можна використовувати різні формати та джерела даних.

Microsoft Azure ML Studio працює з наступними джерелами даних:

- Azure Blob Storage;
- Azure SQL DB;
- Azure SQL DW;
- Azure Table;
- Desktop Direct Upload;
- Hadoop Hive Query;
- Manual Data Entry;
- OData Feed;
- On-prem SQL Server;

- Web URL (HTTP).

Microsoft Azure ML Studio підтримує такі формати даних:

- ARFF;
- CSV;
- SVMLight;
- TSV;
- Excel;
- ZIP.

Після імпорту даних їх можна візуально проаналізувати за допомогою модуля візуалізації, що підтримує:

- точкові діаграми;
- стовпчикові діаграми;
- коробкові графіки;
- гістограми;
- бібліотеки R та Python;
- REPL з Jupyter Notebook;
- змішані матриці;
- дерева рішень.

На етапі попередньої обробки даних необхідно опрацювати дані вибірки. Прості показники статистики, такі як розподіл значень, середнє значення/медіана, відсоток пропущених значень та діапазон значень кожної характеристики допоможуть отримати уявлення про бізнес, а їх аналіз корисний для контролю цілісності даних. Як правило, для цього підходять методи візуалізації.

Слід перевірити дані на предмет інтерпретації (наприклад, щоб впевнитись, що 0 означає нульові, а не пропущені значення). Також на даному етапі підтверджується, що збір даних був здійснений на основі заданих критеріїв і всі аспекти даних зрозумілі, включаючи випадкові зміни даних.

Більшість фінансових даних відрізняються відсутністю деяких значень, чи, навпаки наявністю значень, які не є цільовими для характеристики. Це можуть бути поля, значення яких не були зафіксовані, які більше не використовуються, які були недоступні чи більше не заповнюються, а також неправильно заповнені, викиди або дані, що істотно виділяються.

Якщо деякі статистичні методи, наприклад дерева рішень, індиферентні до пропущених значень, то логістична регресія вимагає, щоб набори даних про клієнтів були повними, без пропущених значень.

Дану ситуацію можна виправити наступними способами:

- виключити всі дані з пропущеними значеннями, проте у випадку фінансових даних в цілому такий аналіз дасть дуже мало даних для обробки;
- виключити з моделі характеристики чи записи, для яких доля пропущених значень істотна (наприклад, більша 50%), особливо, якщо очікується, що в майбутньому рівень пропущених значень буде зберігатись;
- задати пропущені значення статистичними методами.

Дані варіанти припускають, що в пропущених даних відсутні значення, тобто ніякої додаткової інформації аналіз таких даних принести не може. Це не обов'язково так: пропущені значення можуть бути частиною тренду, бути пов'язаними з іншими характеристиками чи бути індикатором «поганої» поведінки клієнта.

Пропущені значення зазвичай невинуваті. Наприклад, клієнт, який нещодавно влаштувався на роботу, скоріш за все залишить поле в анкеті «Стаж на поточному місці роботи (років)» незаповненим. Якщо виключити з аналізу характеристики чи записи з пропущеними значеннями, то аналіз не буде відображати реальної ситуації. Через це рекомендується включити в аналіз пропущені значення. Тобто в цьому методі визнається, що відсутність даних має деяку інформаційну цінність і включення їх в аналіз принесе деяку

інформаційну цінність для бізнесу. Принаймні, пропущені значення слід спочатку проаналізувати, а потім, якщо з'ясується, що вони випадкові та нейтральні з точки зору поведінки, їх можна виключити чи присвоїти їм значення.

Деякі програмні продукти містять алгоритми присвоєння значень пропущеним даним. Для скорингового аналізу рекомендуються методи підстановки, враховуючи значення інших характеристик чи записів. Присвоєння «найбільш часто використаних» чи середніх значень пропущеним полям може призвести до «сплеску» в даних, і диференціація між даними з присвоєними середніми значеннями і тими, які в дійсності мали такі значення, буде неможлива. В результаті може бути втрачена важлива для бізнесу інформація.

«Викиди» - це значення, які випадають з нормального діапазону тієї чи іншої характеристики. Наприклад, розподіл по віку може показати, що вся популяція знаходиться в діапазоні від 18 до 55 років, і декілька випадків – 99,112 та 134 роки. Хоча це і може бути правдою, але, ймовірніше за все. Це результат помилкового вводу даних персоналом. Дані цифри можуть негативно вплинути на результати регресії, і, як правило їх виключають. В деяких випадках замість них можуть бути задані середні значення, тому що зазвичай кількість таких даних не суттєва і не впливає негативно на результат. Проте, спочатку необхідно вивчити викиди, так як вони можуть вказувати на такі проблеми, як шахрайство.

В студії машинного навчання Azure доступні такі функції для попередньої обробки даних:

- очистка пропущених значень;
- додавання стовпця за допомогою виразу;
- додавання відсутніх значень;
- отримання стовпця за зразком;
- фільтрація;

- призначені для користувача перетворення Python, тощо.

На етапі поділу даних на вибірки на навчальну, тестову та перевіірочну.

Навчальна вибірка - вибірка, по якій проводиться налаштування (оптимізація параметрів) моделі. Якщо модель побудована на основі навчальної вибірки, то оцінка якості цієї моделі, виконана за тією ж вибіркою є, як правило, оптимістично зміщена. Таке небажане явище називають перенавчанням. На практиці воно зустрічається дуже часто. Хорошу емпіричну оцінку якості побудованої моделі дає її перевірка на незалежних дані, які не використовувались для навчання.

Перевіірочна (або контрольна) вибірка - вибірка, по якій оцінюється якість побудованої моделі. Якщо навчальна та тестова вибірка незалежні, то оцінка, виконана за тестовою вибіркою, є незміщеною.

Оцінку якості, зроблену на тестовій вибірці, можна застосувати для вибору найкращої моделі.

В студії машинного навчання Azure представлені наступні алгоритми:

1. Двохкласова класифікація

- логістична регресія;
- ліс рішень;
- джунглі рішень;
- дерево рішень, що збільшується;
- нейронна мережа;
- метод опорних векторів;
- Bayes Point Machine;

2. Багатокласова класифікація

- логістична регресія;
- ліс рішень;
- джунглі рішень;
- нейронна мережа;
- one-v-all

3. Регресія

- лінійна;
- Баєсівська;
- Квантильна регресія швидкого лісу;
- регресія Пуасона;
- порядкова

4. Виявлення аномалій

- метод опорних векторів;
- метод К-середніх

Навчання моделі відбувається на навчальній вибірці, а оцінка - на перевірочній.

3.4 Використання отриманих результатів моделі для впровадження маркетингової стратегії в банку

Після опрацювання даних сегменту клієнтів, формується вибірка клієнтів у яких низький рівень дефолту. Дані клієнти будуть цільовим сегментом для майбутньої маркетингової кампанії, що реалізується за допомогою CRM-системи.

Схема маркетингової кампанії наведена на рис.3.3.

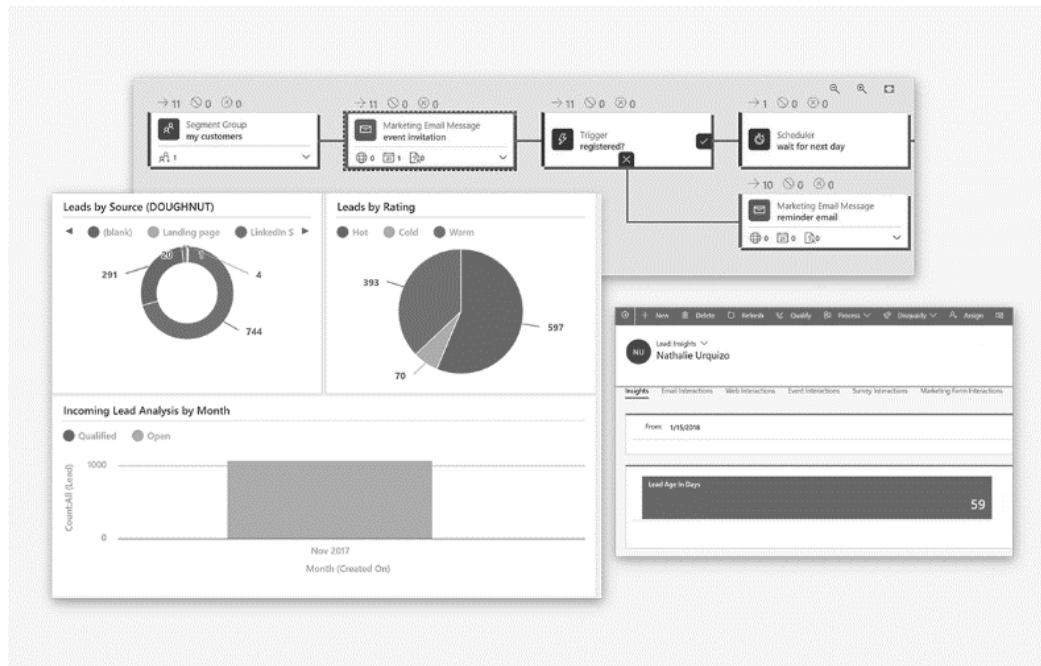


Рисунок 3.3 – Схема маркетингової кампанії

В подальшому маркетингова кампанія може бути реалізована за допомогою наступних каналів комунікації:

- sms-розсилка;
- e-mail розсилка;
- viber розсилка;
- реклама на радіо;
- реклама на телебаченні;
- IVR;
- вихідні дзвінки.

- Висновки до розділу

В даному розділі були сформульовані остаточні вимоги до майбутньої системи підтримки прийняття рішень. Основною задачею системи є узгодження маркетингової політики з політикою видачі кредиту банку.

Головні вимоги з точки зору кінцевих користувачів – простота у використанні та інтуїтивний (зрозумілий) інтерфейс.

Основними цілями СППР визначені:

- підвищення лояльності та довіри клієнтів банку;
- зниження витрат на маркетингові кампанії;
- зменшення кількості відмов у наданні кредитів;
- зменшення кількості заборгованості ;
- збільшення операційної ефективності .

Також була розроблена архітектура програми та вибрана платформа для проведення досліджень.

РОЗДІЛ 4 ПОБУДОВА СКОРИНГОВИХ МОДЕЛЕЙ ТА АНАЛІЗ РЕЗУЛЬТАТІВ

4.1 Трансформація, очистка та опис отриманих даних

Після формування вибірки клієнтів банку з внутрішньої бази отримуємо файл `mc_cred_scor_v1 (.csv)`.

Нижче в табл.4.1. описано дані після обробки, що використовувались при побудові скорингової моделі.

Таблиця 4.1 – Опис даних після первинної обробки

Назва змінної	Тип значень	Опис/значення
age	Номінальний	Вік клієнта у роках
marital_status	Номінальний	Стать та сімейний стан 1: чоловік – розлучений 2: жінка – одружена 3: чоловік – неодружений 4: чоловік – одружений 5: жінка – незаміжня 6: жінка – розлучена
resident	Номінальний	1: резиденти 2: отримав громадянство до 3 років тому 3: отримав громадянство від 3 до 7 років тому 4: отримав громадянство більше ніж 7 років тому
is_phone	Бінарний	1: немає даних 2: вказаний
job_status	Номінальний	1: безробітний 2: знаходиться на обліку в центрі зайнятості 3: працевлаштований 4: підприємець

Продовження табл. 4.1

term_empl	Номінальний	Стаж роботи 1: безробітний 2: ... < 1 року 3: 1 <= ... < 4 роки 4: 4 <= ... < 7 років 5: ... >= 7 років
foreign_worker	Бінарний	Влаштований закордоном 1: Так 2: Ні
registration_type	Номінальний	Тип реєстрації 1: орендоване помешкання 2: власне помешкання 3: інше (помешкання батьків, тощо)
property	Номінальний	1: Нерухомість та автомобіль 2: Нерухомість 3: Автомобіль 4: Не визначено
cur_income		Сума доходу 1: 0 < ... < 5000 2: 5000 <= ... < 12000 3: 12000 >= 20000 4: >20000
Credit_amount	Номінальний	Сума кредиту
cred_term_m	Номінальний	Термін кредиту в місяцях
credit_product	Номінальний	Види кредитних продуктів 0: кредит на новий автомобіль 1: кредит на вживаний автомобіль 2: кредит на ремонт 3: кредит на побутову техніку 4: споживчий кредит 5: кредит під заставу депозиту 6: кредит на навчання 7: кредит готівкою 8: кредит на нерухомість 9: кредит під заставу депозиту 10: інше

Продовження табл. 4.1

cred_history	Номинальний	Кредитна історія		
		0: не було кредитів 1: всі минулі кредити оплачені вчасно 2: наявні кредити в даному банку 3: наявні кредити в іншому банку 4: наявні кредити в даному та іншому банку		
rate_fr_income	Номинальний	Сума	відсотків	відносно доходу
other_debtors	Номинальний	Наявність поручителя/гаранта		
		1: немає 2: поручитель 3: гарант		
depends	Номинальний	Кількість членів родини		
num_exist_cred	Номинальний	Кількість активних кредитів		
dep_amount	Номинальний	Сума депозитів		
		1: до 1000 2: 1000-5000 3: 5000-15000 4: 15000- 20000 5: більше 20000		
education	Номинальний	Освіта		
		1: вища повна 2: середня 3: вища неповна		
good_bad	Бінарний	Цільова	змінна,	показник дефолту:
		0: дефолт 1: не дефолт		

На першому етапі обробки вхідних даних була проведена очистка повних дублікатів даних та заповнення пропусків даних для наступних характеристик за встановленими правилами:

- is_phone – 1 (немає даних);

- job_status – 1 (безробітний);
- term_empl : якщо не вказано job_status – 1 (безробітний), якщо вказано – 2 (< 1 року);
- foreign_worker – 2 (Ні);
- registration_type – 1 (орендоване помешкання)
- property – 4 (Не визначено);
- other_debtors – 1 (немає).

Також були видалені рядки з аномальними даними:

- Credit_amount;
- Birth_date;
- cur_income;
- credit_product.

На етапі трансформації для наступних показників було здійснено групування та узагальнення:

- marital_status;
- resident;
- term_empl;
- cur_income;

Показник age був розрахований на основі даних про дату народження клієнта.

В результаті обробки було отримано таблицю з 2000 записів та 21 колонкою. Далі представлено детальний опис вибірки даних.

Розподіл значень доходу клієнта представлений на рис.4.1. Найбільше значення 39,4% припадає на варіант 4; значення медіани дорівнює 2; пропусків немає.

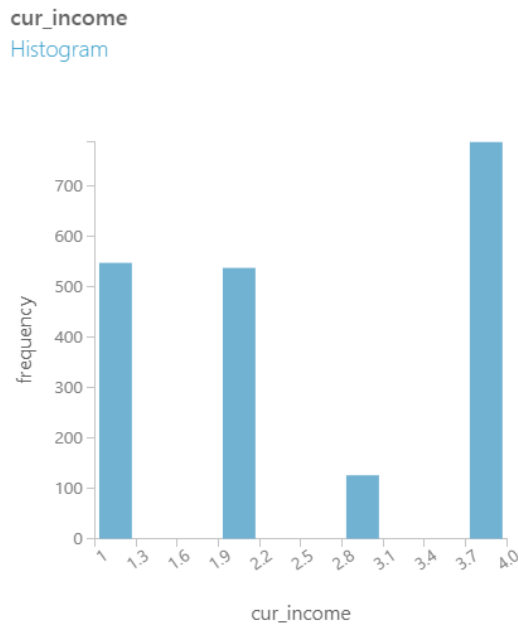


Рисунок 4.1 - Розподіл значень доходу клієнта

Розподіл значень терміну кредиту представлений на рис. 4.2. Мінімальне значення дорівнює 4 місяцям, максимальне 72 місяці (6 років). Середнє значення – 18 місяців. Пропусків даних відносно даного параметру немає.

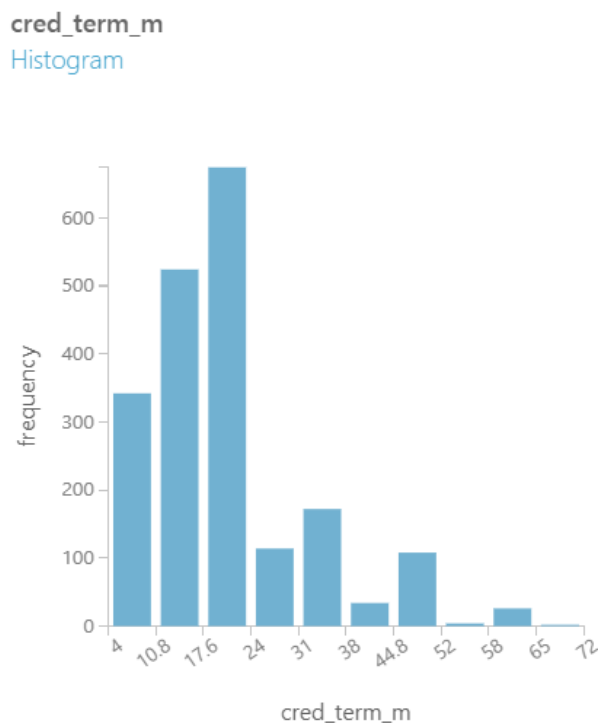


Рисунок 4.2 - Розподіл значень терміну кредиту

Розподіл значень агрегованого показника кредитної історії представлений на рис.4.3, найбільше значення 53% припадає на варіант 2;

значення медіани дорівнює 2. Пропусків даних відносно даного параметру немає.

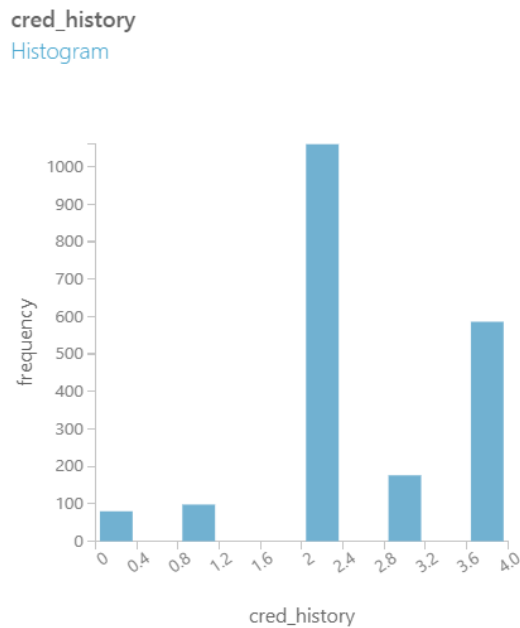


Рисунок 4.3 - Розподіл значень агрегованого показника кредитної історії

Розподіл значень показника суми кредиту представлений на рис.4.4, Мінімальна сума кредиту становить 250, максимальна – 18424.Пропусків даних відносно даного параметру немає.

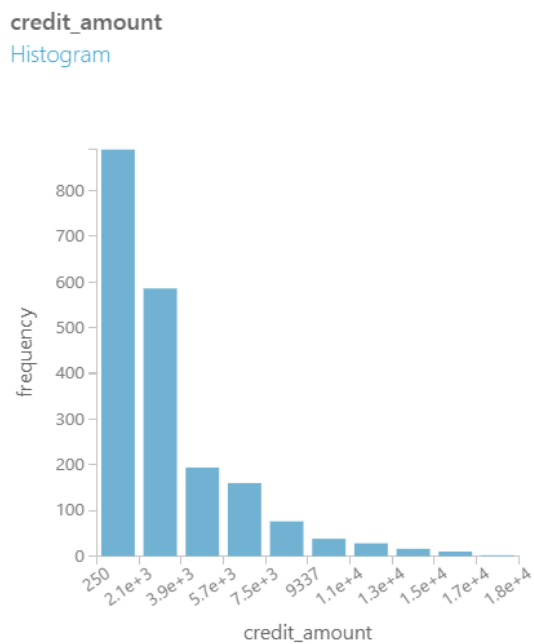


Рисунок 4.4 - Розподіл значень показника суми кредиту

Розподіл значень показника суми депозиту представлений на рис.4.5, 60,3 % припадає на допозити до 1000 . Пропусків даних відносно даного параметру немає.

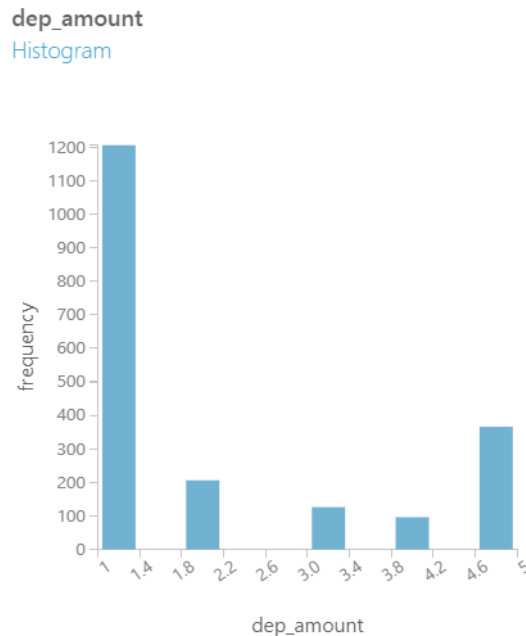


Рисунок 4.5 – Розподіл значень показника суми депозиту

Розподіл значень показника стажу роботи представлений на рис.4.6, 34 % припадає на стаж 3х років. Пропусків даних відносно даного параметру немає.

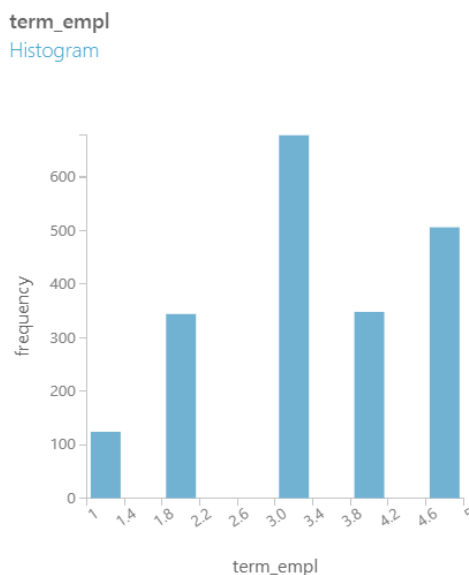


Рисунок 4.6 – Розподіл значень показника стажу роботи

Розподіл значень показника освіти представлений на рис.4.7, 90,7% припадає на вищу повну освіту. Пропусків даних відносно даного параметру немає.

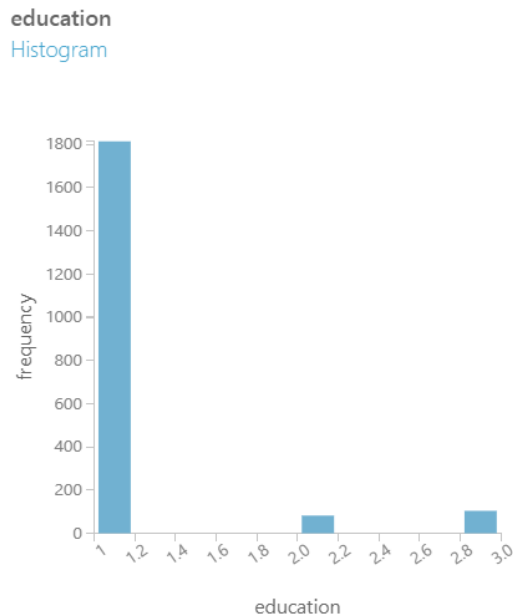


Рисунок 4.7 – Розподіл значень показника освіти

Розподіл значень показника resident представлений на рис.4.8 , 41,3% припадає на позичальників, що стали громадянами більше 4 років тому. Пропусків даних відносно даного параметру немає.

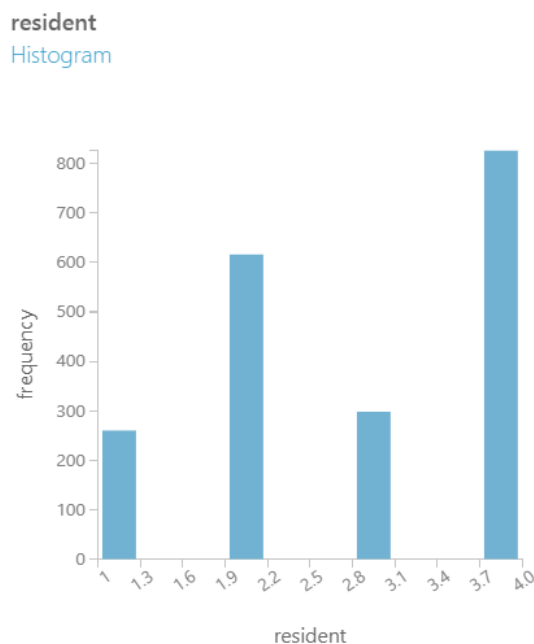


Рисунок 4.8 – Розподіл значень показника resident

Розподіл значень показника власності представлений на рис.4.9, Найбільше значення 33,2% позичальників мають власне авто. Пропусків даних відносно даного параметру немає.

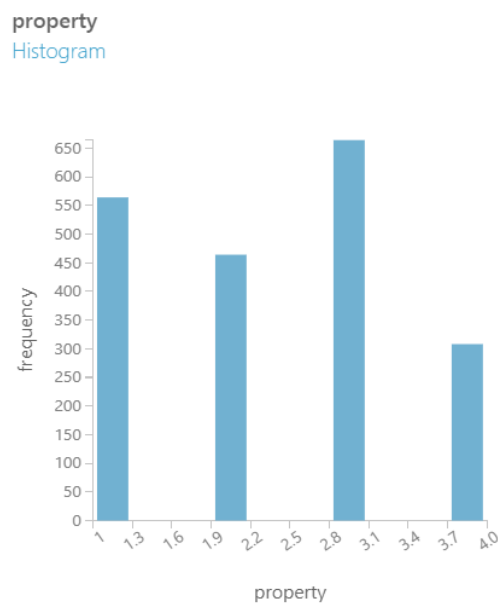


Рисунок 4.9 – Розподіл значень показника власності

Розподіл значень показника віку позичальників представлений на рис.4.10. Медіана дорівнює 33 роки, найменше значення – 19 років.Пропусків даних відносно даного параметру немає.

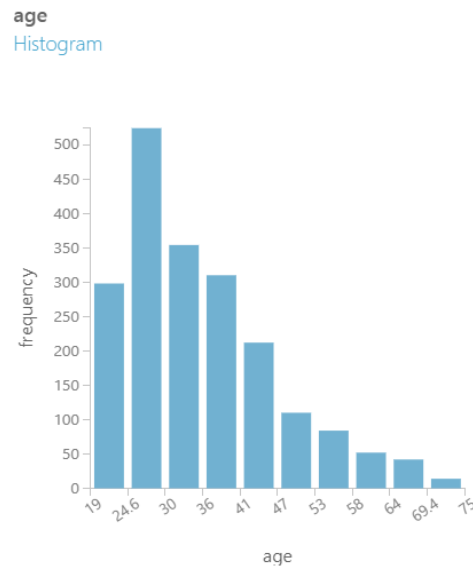


Рисунок 4.10 – Розподіл значень показника віку позичальників

Розподіл значень показника реєстрації позичальників представлений на рис. 4.11. 71,3 % позичальників мають власне поомешкання. Пропусків даних відносно даного параметру немає.

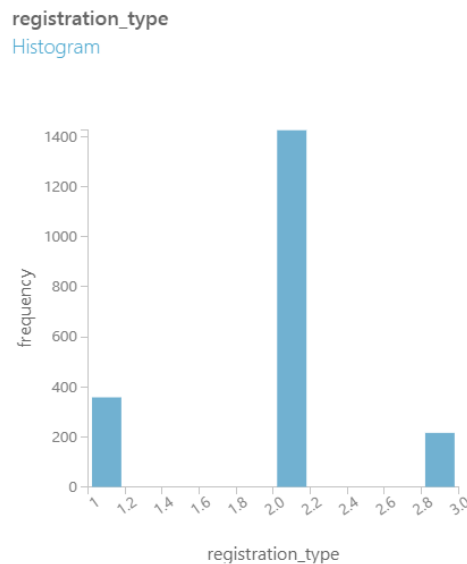


Рисунок 4.11 – Розподіл значень показника реєстрації позичальників

Розподіл значень кількості кредитів позичальників представлений на рис.4.12. 63,3 % позичальників мають один кредит. Пропусків даних відносно даного параметру немає.

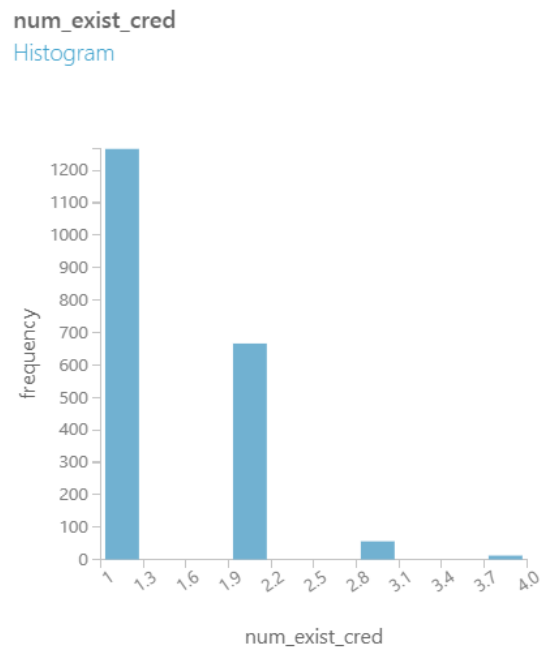


Рисунок 4.12 – Розподіл значень кількості кредитів позичальників

4.2 Проведення експериментів для визначення найкращої скорингової моделі

4.2.1 Логістична регресія

Модель прогнозування кредитоспроможності позичальників з використанням алгоритму логістичної регресії представлена на рис.4.13

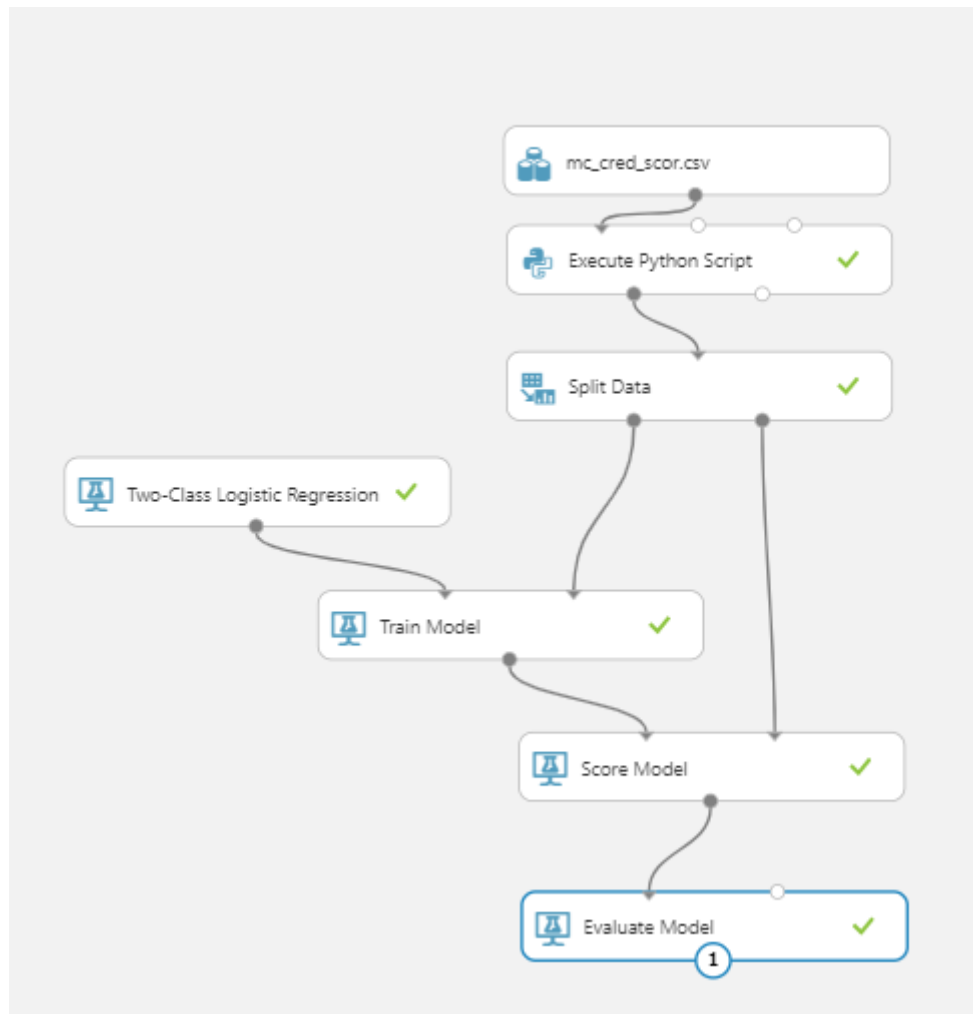


Рисунок 4.13 - Модель прогнозування кредитоспроможності на основі алгоритму логістичної регресії

На рис. 4.14 представлені параметри, що використовувались при налаштуванні моделі.

Two-Class Logistic Regression

Create trainer mode

Single Parameter ▼

Optimization tolerance

1E-07

L1 regularization weight

1

L2 regularization weight

1

Memory size for L-BFGS

20

Random number seed

Рисунок 4.14 – Параметри для алгоритму логістичної регресії

Графік ROC-кривої наведено на рис.4.15.

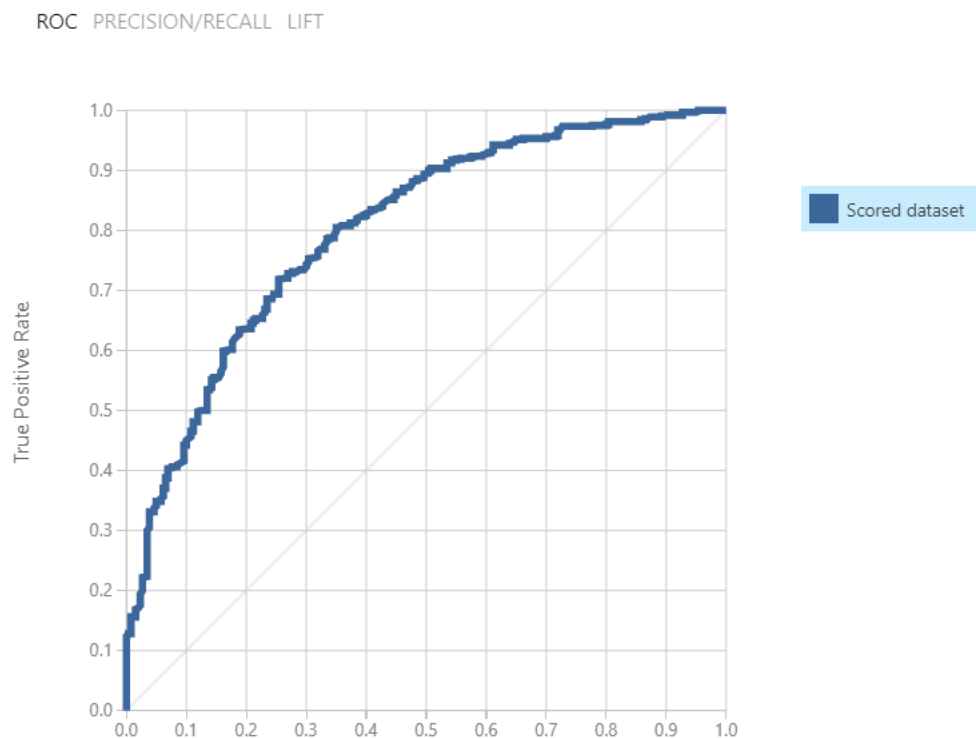


Рисунок 4.15 – ROC- крива для алгоритму логістичної регресії

Для ваги регуляризації L1 та ваги регуляризації L2 були введені значення для параметрів регуляризації L1 і L2. Дані значення не мають бути рівні нулю.

Регуляризація - це метод запобігання перенавчання, коли моделі з екстремальними значеннями коефіцієнтів нараховуються штрафи. Регуляризація працює шляхом додавання штрафів. Таким чином, точна модель з екстремальними значеннями коефіцієнта буде більше оштрафована.

Регуляризація L1 і L2 має різні ефекти та використання. L1 можна застосувати до розріджених моделей, що є корисним при роботі з великорозмірними даними. Цей алгоритм підтримує лінійну комбінацію значень регуляризації L1 і L2, тобто якщо $x = L1$ та $y = L2$, то $ax + by = c$ визначає лінійний пропуск термінів регуляризації.

Результати проведеного експерименту наведені у табл.4.2

Таблиця 4.2. – Результати експерименту з використанням логістичної регресії

Показник	Значення
True Positive (TP)	570
True Negative (TN)	131
False Positive (FP)	129
False Negative (FN)	70
Accuracy	0,779
TPR (Recall)	0,890
TNR	0,503
FPR	0,496
FNR	0,109
Precision (PREC)	0,815
MCC	0,429
F1	0,851
AUC	0,802

Даний алгоритм показав середні результати. Не дивлячись на досить непогану точність моделі – 0,779, показник неправильно розпізнаних «хороших» позичальників є високим.

Даний показник є дуже важливим, так як слід розуміти, що потенційні клієнти не потрапили до сегменту маркетингової кампанії, а як результат банк може недоотримати дохід від запропонованого продукту.

Також досить великим є показник нерозпізнаних дефолтів, що суттєво збільшує кредитні ризики банку. Показник кореляції свідчить про відсутність проблеми перенавчання модулі.

4.2.2 Древа рішень

Модель прогнозування кредитоспроможності позичальників з використанням алгоритму дерев рішень представлена на рис.4.16

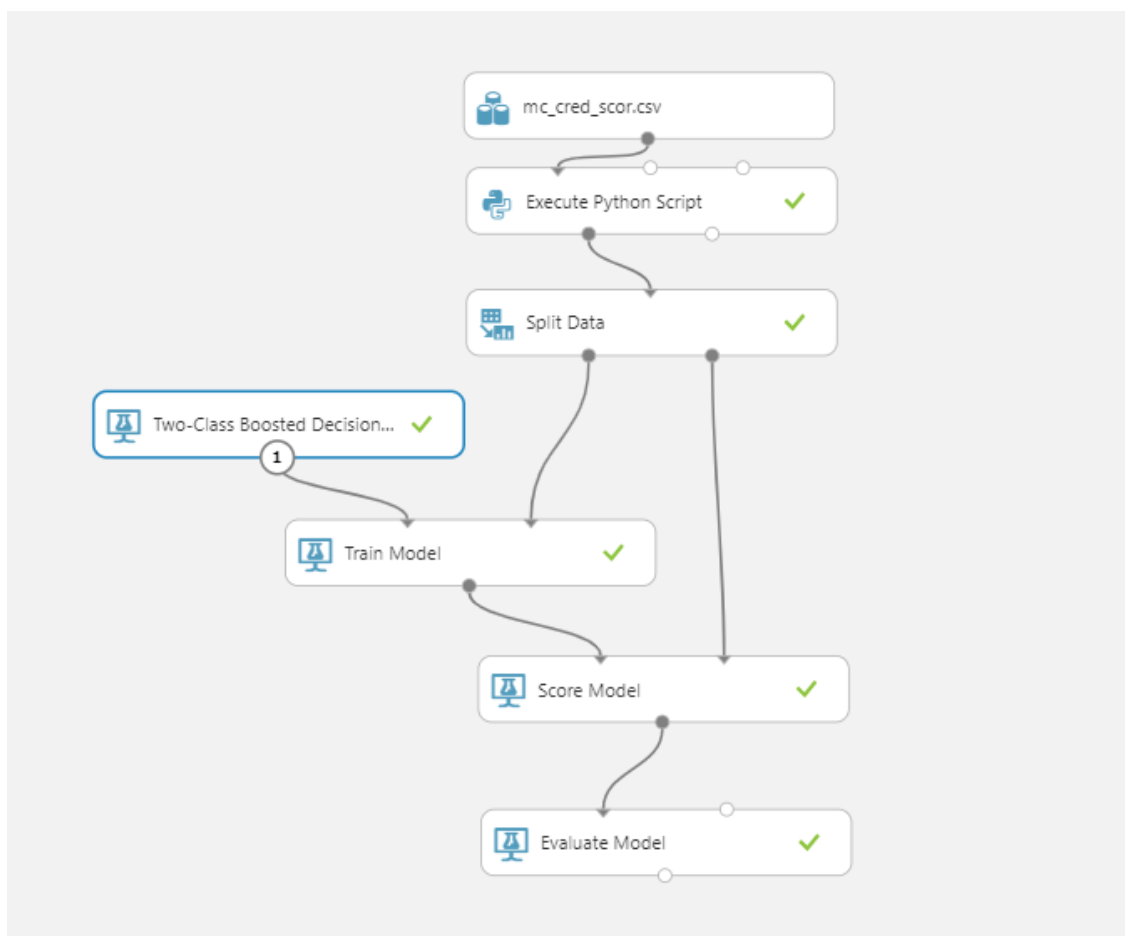



Рисунок 4.16 - Модель прогнозування кредитоспроможності на основі алгоритму дерев рішень

На рис. 4.17 представлені параметри, що використовувались при налаштуванні моделі.


▲ Two-Class Boosted Decision Tree

Create trainer mode


Single Parameter ▼

Maximum number of leaves per tree 


20

Minimum number of samples per leaf node 

10

Learning rate 

0.2

Number of trees constructed 

100

Рисунок 4.17 – Параметри для алгоритму дерев рішень

Максимальна кількість листків на дерево була визначена стандартним значенням 20, мінімальна – 10.

Графік ROC-кривої наведено на рис.4.18.

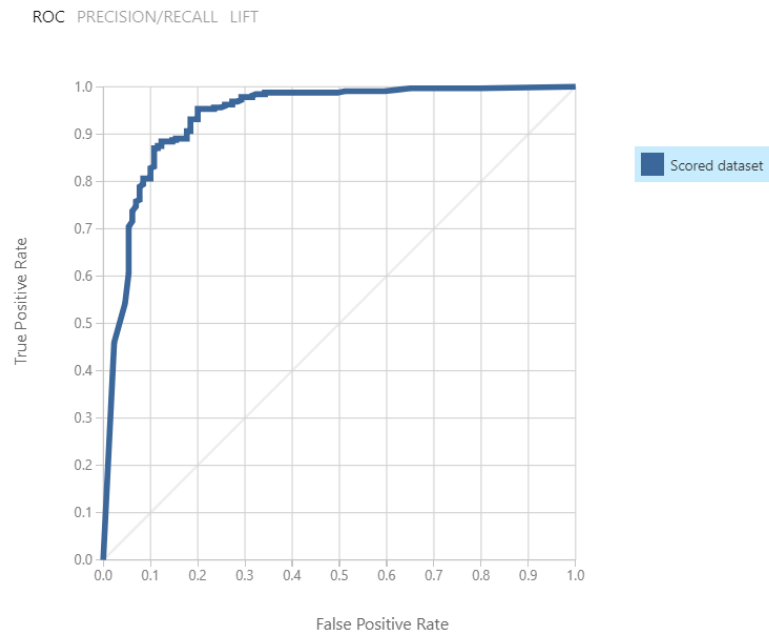


Рисунок 4.18 – ROC- крива для алгоритму дерев рішень

Результати проведеного експерименту наведені у табл.4.3

Таблиця 4.3 – Результати експерименті з використанням логістичної регресії

Показник	Значення
True Positive (TP)	592
True Negative (TN)	212
False Positive (FP)	48
False Negative (FN)	48
Accuracy	0,893
TPR (Recall)	0,925
TNR	0,815
FPR	0,185
FNR	0,075
Precision (PREC)	0,925
MCC	0,151
F1	0,925
AUC	0,940

Алгоритм дерев рішень показав досить непогані результати.

Неправильно розпізнаних клієнтів - 48, що становить лише 5 % від всієї

вибірки. Також 5 % становлять неправильно розпізнані дефолти. Даний показник допустимий, проте не є ідеальним.

4.3 Порівняння отриманих результатів

Після проведення окремих експериментів з алгоритмами логістичної регресії та дерева рішень переходимо до порівняння двох моделей для визначення тієї, що ляже в основу системи підтримки прийняття рішень для впровадження маркетингових стратегій у банку.

Графіки ROC-кривих обох моделей представлені на рис.4.19.

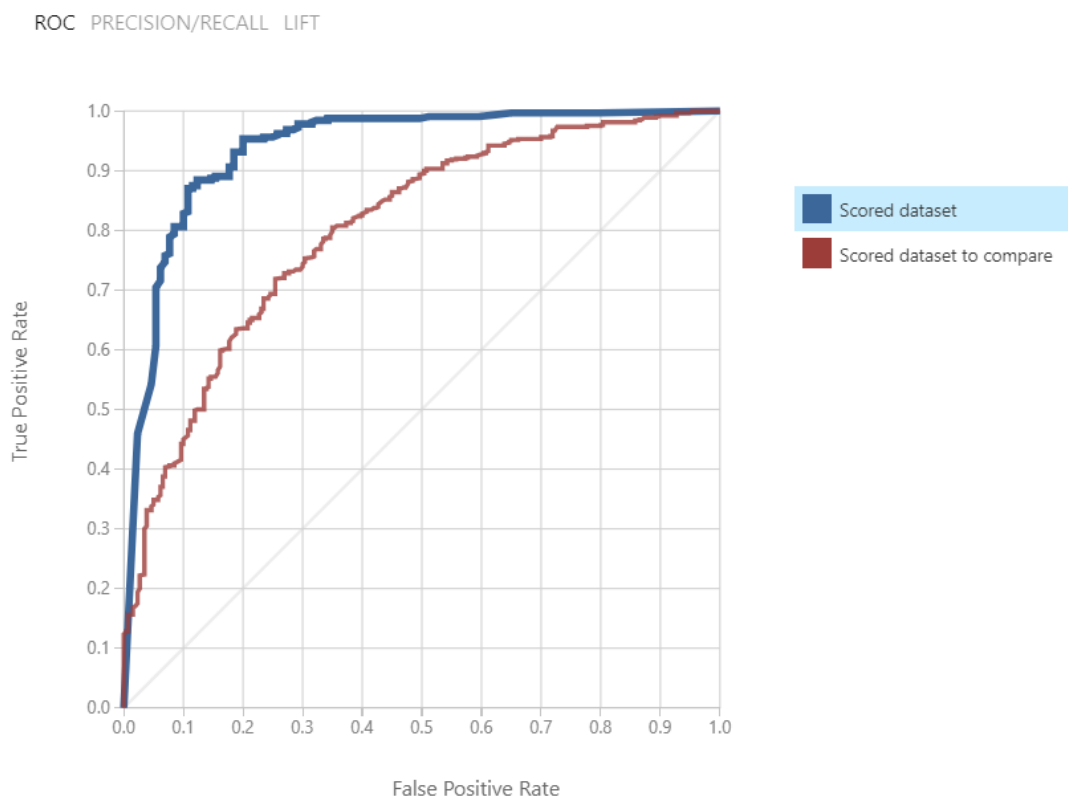


Рисунок 4.19 –Порівняльний аналіз ROC-кривих скорингових моделей

Як видно з рисунку, модель з використанням алгоритму дерев рішень показала набагато кращі результати для даної вибірки. Детальний порівняльний аналіз здійснено за допомогою значень табл.4.4.

Таблиця 4.4 – Порівняльна таблиця результатів попередніх досліджень

Показник	Логістична регресія	Дерева рішень
True Positive (TP)	570	592
True Negative (TN)	131	212
False Positive (FP)	129	48
False Negative (FN)	70	48
Accuracy	0,779	0,893
TPR (Recall)	0,890	0,925
TNR	0,503	0,815
FPR	0,496	0,185
FNR	0,109	0,075
Precision (PREC)	0,815	0,925
MCC	0,429	0,151
F1	0,851	0,925
AUC	0,802	0,940

Алгоритм дерев рішень показав на 3,85% кращий результат розпізнавання «хороших» клієнтів, та на 62,7% - дефолтів.

Тобто, можна зробити висновок, що даний алгоритм суттєво зменшує ризик для банку та збільшує сегмент «хороших» клієнтів для маркетингової кампанії.

Точність моделі на 11,4 пункти більша у алгоритму дерева рішень. Суттєво відрізняється показник кореляції, що свідчить про відсутність перенавчання моделей.

Висновки до розділу

В даному розділі описані етапи первинної обробки даних та підготовки до побудови скоригової моделі. Для обробки даних були застосовані наступні методи:

- видалення дублікатів записів;
- видалення записів, що містять викиди;
- обробка пропусків даних;
- групування характеристик.

Детально продемонстровані вибірки даних після обробки.

Підготовлена вибірка даних була використана в навчанні та перевірці моделей з використанням алгоритмів логістичної регресії та дерев рішень. Після обробки результатів та розрахунку показників оцінки скорингових моделей виявлено, що для вирішення поставлених задач найкраще підходить алгоритм дерев рішень.

Алгоритм дерев рішень показав на 3,85% кращий результат розпізнавання «хороших» клієнтів, та на 62,7% - дефолтів.

Тобто, можна зробити висновок, що даний алгоритм суттєво зменшує ризик для банку та збільшує сегмент «хороших» клієнтів для маркетингової кампанії.

РОЗДІЛ 5 АНАЛІЗ РИНОВИХ МОЖЛИВОСТЕЙ СТАРТАП-ПРОЕКТУ

5.1 Опис ідеї проекту

Опис ідеї стартап проекту описано в таблиці 5.1.

Таблиця 5.1 - Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Система підтримки прийняття рішень допомагає співробітнику відділу маркетингу банку у впровадженні маркетингової стратегії шляхом оцінки сегменту клієнтів на кредитоспроможність та використання даного сегмента в маркетинговій кампанії.	Оцінка вибраного сегменту для продукту, що є об'єктом маркетингової кампанії	Допомога у виборі сегменту
	Скоринг клієнта перед відправкою маркетингової (акційної) пропозиції	Зменшення ризику пропозиції клієнту недоступного продукту (Підвищення лояльності клієнтів)
	Виключення із сегменту потенційних неплатоспроможних (для даного продукту) клієнтів	Зменшення витрат на маркетинг

До сильних сторін можна віднести:

- можливість оптимізації витрат на маркетинг;
- зменшення ризику дефолту;
- підвищення рівня лояльності клієнта.

Детальна характеристика сильних, слабких та нейтральних сторін ідеї проекту зображено в таблиці 5.2.

Таблиця 5.2 - Визначення характеристик ідеї проекту

№ п/п	Техніко- економічні характеристики ідеї	Потенційні) товари/концепції конкурентів			W (слабка сторона)	N (нейтр. сторона)	S (сильна сторона)
		Мій про ект	FIC O Scori ng soluti ons	SAS Credit Factor y			
1.	Оптимізація витрат на маркетинг	+	-	-			+
2.	Зменшення ризику дефолту	+	+	+		+	
3.	Підвищення рівня лояльності клієнта	+	-	+			+
5.	Оптимізація процесу впровадження маркетингової стратегії	+	+	+		+	
5.	Відомість бренду	-	+	+	+		
6.	Наявність бази кредитних історій	-	+	-			+

5.2 Технологічний аудит ідеї проекту

Технологічний аудит ідеї проекту наведений у таблиці 5.3.

Таблиця 5.3 - Технологічна здійсненність ідеї проекту

№ п/п	Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
1	Інтерфейс програми для наземних версій	C#	Розробка інтерфейсу	Технологія загальнодоступна
2	Веб-кабінет управління	HTML, JS, CSS	Розробка веб-кабінету	Технологія загальнодоступна
3	Моделювання та прогнозування	Azure ML Studio/R	Необхідно реалізувати моделі оцінки кредитоспроможності	Технологія загальнодоступна
	Конектор до CRM-системи	Веб-сервіс	Доступна публікація веб-сервісу	Технологія загальнодоступна
Обрана технологія реалізації ідеї проекту: для реалізації проекту обрана мова програмування C#.				

5.3 Аналіз ринкових можливостей запуску стартап-проекту

Характеристика потенційного ринку стартап-проекту наведена у таблиці 5.4.

Таблиця 5.4 - Попередня характеристика потенційного ринку стартап-проекту

№ п/п	Показники стану ринку (найменування)	Характеристика
1	Кількість головних гравців, од	3
2	Загальний обсяг продаж, грн/ум.од	3 000 000
3	Динаміка ринку (якісна оцінка)	Стагнує
4	Наявність обмежень для входу (вказати характер обмежень)	Висока точність розпізнавання
5	Специфічні вимоги до стандартизації та сертифікації	Відповідність до законодавства
6	Середня норма рентабельності в галузі (або по ринку), %	60

Характеристика потенційних клієнтів стартап-проекту наведена в таблиці 5.5.

Таблиця 5.5 - Характеристика потенційних клієнтів стартап-проекту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1	Таргетинг цільових аудиторій з додатковим скорингом	Банки та кредитні спілки, кредитні кафе	Різний об'єм вхідних даних	Продаж послуг має збільшитись мінімум на 15%

До можливих загроз на початковому етапі можна віднести :

- відсутність попиту;

- неточне визначення дефолту споживача за рахунок обмежених даних.

Можливі загрози для стартап-проекту наведені у таблиці 5.6.

Таблиця 5.6 - Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1	Відсутність попиту	Бізнес може не оцінити переваги продукту, або ж у цілому відмовитися від зміни бізнес-процесів у маркетинговому відділі	Навести дані про ефективність продукту, запропонувати знижку або безкоштовну демо версію потенційному клієнту.
2	Неточне визначення дефолту споживача послуг	Параметри моделі не є універсальними для всіх клієнтів	Запропонувати кастомізацію продукту під певного клієнта.

Фактори можливостей наведені у таблиці 5.7.

Таблиця 5.7 - Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1	Продаж готового рішення	Продаж готового рішення у великій компанії сфери аналітики	Виділення частини штату на реалізацію проекту.

Для аналізу конкуренції був проведений детальний ступеневий аналіз на ринку.

Результати ступеневого аналізу конкуренції на ринку зображено у таблиці 5.8.

Таблиця 5.8 - Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1. Вказати тип конкуренції - олігополія	Присутня невелика кількість фірм. Більшість ринку контролюють фірми-гіганти	Більш вигідні умови на тендерах, агресивний маркетинг, направленість на задоволення потреб клієнта
2. Регіональна конкуренція	Гравці ринку – інтернаціональні підприємства	Вихід на ті ринки, які ще не зайняті конкурентами
3. Внутрішньогалузева конкуренція	Гравці ринку знаходяться в одній галузі	Направленість на задоволення потреб клієнта
4. Товарно-видова конкуренція	Усі продукти гравців ринку мають одне призначення	Направленість на задоволення потреб клієнта
5. Конкурентні переваги нецінові	Продукти відрізняються гнучкістю, функціоналом (незначно) і надійністю.	У маркетингу неявно порівнювати власний продукт з іншими, робити вигідні цінові

		пропозиції, наголошувати на якості обслуговування
6. Марочна конкуренція	Значна увага приділяється бренду, що розробив продукт	Продаж готового рішення для аналітичних систем відомих брендів

Проведений аналіз конкуренції в галузі зображено у таблиці 5.9.

Таблиця 5.9 - Аналіз конкуренції в галузі за М. Портером

	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари- замінники
Складові аналізу	FICO, Microsoft, SAS	Відсутні референси	Відсутній вплив	Споживачі замовляють додаткові модулі до програмного продукту	SAS Credit Factory
Висновки	Продукти конкурують в різних сегментах ринку	Можливість виходу на ринок є через партнерську мережу	Постачальники не диктують умови роботи на ринку.	Клієнти диктують умови лише для додаткових модулів, що розроблюють для них.	Для того, щоб потенційний клієнт обрав саме цей продукт, необхідно, щоб він був максимально простий, швидкий і надійний у використанні, а також був гнучким і масштабованим.

Відповідно до ідеї були визначені наступні фактори
конкурентоспроможності:

- інтеграція,
- модульність,
- гнучкість.

Фактори конкурентоспроможності та їх обґрунтування наведені в таблиці 5.10.

Таблиця 5.10 - Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
1	Інтеграція	Наявні порти для найпопулярніших CRM-систем
2	Модульність	Кожен замовник індивідуально обирає моделі для себе, при цьому не втрачаючи можливість докупити інші пакети в майбутньому
3	Гнучкість	Кожен замовник має можливість замовити розширення функціоналу продукту під його конкретні задачі

Порівняльний аналіз сильних та слабких сторін проекту відображено у таблиці 5.11.

Таблиця 5.11 - Порівняльний аналіз сильних та слабких сторін RFS

№ п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з BA Consulting						
			-3	-2	-1	0	+1	+2	+3
1	Інтеграція						*		
2	Модульність				*				
3	Гнучкість				*				
4	Розділення осіб				*				

Важливим елементом аналізу стартап-проекту є SWOT-аналіз, що являє собою аналіз сильних, слабких сторін проекту, а також можливостей та загроз.

SWOT-аналіз проекту наведено в таблиці 5.12.

Таблиця 5.12 - SWOT-аналіз стартап-проекту

Сильні сторони: Сегментація потенційних клієнтів у поєднанні зі скорингом кредитоспроможності (використання кредитного скорингу в маркетингових рішеннях)	Слабкі сторони: Неточне визначення дефолту споживача послуг, відсутність референсів
Можливості: Продаж готових рішень, продаж рішень через партнерів	Загрози: Відсутність попиту

Альтернативи ринкового впровадження проекту розглянуто в таблиці 5.13.

Таблиця 5.13 - Альтернативи ринкового впровадження стартап-проекту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1	Сервис onpremise	Висока	12-18 місяців

2	Хмарний сервіс за допомогою рішень Microsoft	Висока	4-5 місяців
3	Хмарний сервіс «своїми силами»	Середня	20-25 місяців

5.4 Розроблення ринкової стратегії проекту

Опис та вибір цільових груп потенційних клієнтів зображено в таблиці 5.14.

Таблиця 5.14 - Вибір цільових груп потенційних споживачів

№ п/п	Опис цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Малий бізнес	Низька готовність, через необхідність вкладання великих коштів.	Низький попит	Середня	Вхід в сегмент складний
2	Середній бізнес	Середня готовність. В залежності від виду бізнесу, готовність різниться.	Середній попит	Вище середньої	Вхід в сегмент достатньо складний

3	Великий бізнес	Абсолютна готовність. Більшість компаній такого рівня застосовували чи планують застосовувати інтелектуальні системи прогнозування, тому повністю готові розглянути альтернативні рішення	Високий попит	Середня	Вхід в сегмент складний
Які цільові групи обрано: 2,3					

В таблиці 5.15 зображено вибір базової стратегії розвитку.

Таблиця 5.15 - Визначення базової стратегії розвитку

№ п/п	Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
	Хмарний сервіс за допомогою рішень Microsoft	Таргетні пропозиції бізнесу, проведення презентації функціональних рішень на ярмарках та конференціях, підхід до обслуговування клієнтів	Інтегрованість з продуктами Microsoft	Наслідкування лідера

Відповідно до визначених конкурентів та їх можливої поведінки була визначена наступна стратегія конкурентної поведінки стартап-проекту, що ґрунтується на стратегії пошуку нових споживачів та переманювання споживачів від конкурентів та акценті на обслуговуванні клієнтів.

В таблиці 5.16 наведено визначення базової стратегії конкурентної поведінки.

Таблиця 5.16 - Визначення базової стратегії конкурентної поведінки

№ п/п	Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки
1	Ні	Можливі обидва варіанти	Стандартні функціональні модулі будуть виконувати схожі функції, що і конкуренти	Унікальна цінова політика, функціональні інновації, підхід до обслуговування клієнта

В таблиці 5.17 наведено визначення стратегії позиціонування.

Таблиця 5.17 - Визначення стратегії позиціонування

№ п/п	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформулювати комплексну позицію власного проекту (три ключових)
	Висока якість прогнозування в клієнтській сфері застосування	Розробка та удосконалення існуючих модулів на основі потреб ринку та інформації від клієнтів	Спеціалізовані рішення, хмарні сервіси	Прогнозування, лояльність, зменшення ризиків

5.5 Розроблення маркетингової програми стартап-проекту

В таблиці 5.18 представлені ключові переваги концепції потенційного товару.

Таблиця 5.18 - Визначення ключових переваг концепції потенційного товару

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1	Простота застосування	Зрозумілий інтерфейс	Забезпечує легке та продуктивне вирішення більшої кількості задач бізнесу
2	Спеціалізовані рішення	Вирішення задач відділу маркетингу	Забезпечує більш ефективне вирішення задач у звуженій сфері застосування
3	Технічні ресурси	Хмарні сервіси	Дозволяє користуватись рішенням за рахунок віддалених технічних потужностей
4	Підвищення лояльності	Зменшення ризиків	Безпосередній вплив на витрати відділу маркетингу

Товар за задумом – це система підтримки прийняття рішень для впровадження маркетингової стратегії шляхом оцінки сегменту клієнтів на кредитоспроможність та використання даного сегмента в маркетингових кампаніях банку. Скорингові моделі високої точності. Надійність роботи у високонавантажених системах. Опис трьох рівнів моделі товару відображено у таблиці 5.19.

Таблиця 5.19 - Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
I. Товар за задумом	Система підтримки прийняття рішень для впровадження маркетингової стратегії шляхом оцінки сегменту клієнтів на кредитоспроможність та використання даного сегмента в маркетингових кампаніях банку.		
II. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	Швидкодія	Нм	Тх/Тл/Е
	Ефективність	Нм	Тх/Тл
	Користувацький інтерфейс	Нм	Е
	Якість: стандарти безпеки, відповідно до законодавства та постанов регуляторів. Скорингові моделі високої точності. Надійність роботи у високонавантажених системах.		
	Пакування: Рішення		
	Марка: ВА Consulting (Banking Analytics Consulting)		
III. Товар із підкріпленням	До продажу: Інтеграція з існуючими системами для автоматизації маркетингових кампаній шляхом використання методів скорингу		
	Після продажу: Технічна підтримка		
Патент. Закритий код. Захищений від можливості декомпіляції.			

Визначення меж встановлення ціни показано в таблиці 5.20.

Таблиця 5.20 - Визначення меж встановлення ціни

№ п/п	Рівень цін на товари- замінники	Рівень цін на товари- аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
1	-	Від 400\$ до 5000\$ + кастомізація залежності від складності проекту	Високий	3000\$ +кастомізація та технічна підтримка

Формування системи збуту зображено в таблиці 5.21.

Таблиця 5.21 - Формування системи збуту

№ п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
1	Таргетні пропозиції для компаній	Презентації функціоналу, безкоштовні пілотні проекти	-	Прямі продажі
2	Завантаження готового рішення з Microsoft Store	-	Обмежується рейтинговою системою Microsoft	Публікація на Microsoft Store

Концепція маркетингових комунікацій відображена у таблиці 5.22.

Таблиця 5.22 - Концепція маркетингових комунікацій

№ п/п	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
1	Середній бізнес – оптимальні рішення за невисоку ціну	Соціальні мережі, таргетні двзінки, конференції.	Автоматизація маркетингових кампаній на основі методів скорингу	Короткий опис переваг продукту, важливість комунікацій лише з тими клієнтами, хто поверне кредит.	Пропозиції кредитів лише надійним клієнтам
2	Великий бізнес – повноцінні рішення для покращення якості маркетингових кампаній	Конференції, демонстрації, презентації бізнес-кейсів, прямі продажі	Автоматизація маркетингових кампаній на основі методів скорингу	Донести ефективність підходу «Рекламна кампанія кредитів лише кредитоспроможним клієнтам»	Автоматизований маркетинг кредитних продуктів лише надійним клієнтам.

Висновки до розділу

Таким чином, виходячи з результатів аналізу відповідного ринку послуг для фінансової галузі України, було досліджено продукт з точки зору його затребуваності та конкурентоспроможності. Виявлено сильні та слабкі сторони продукту. Також, відповідно до проведеної сегментації ринку,

виявлено ключові напрямки маркетингу: середній та великий бізнес. Для цих сегментів розроблено стратегію виводу продукту на ринок, ключові цінності та переваги, які є основою ринкових комунікацій для успішної реалізації розробленої маркетингової стратегії. Варто зауважити, що конкуренція в усіх сегментах ринку є високою, тому існує необхідність регулярного проведення аналізу ринку послуг з можливістю корекції напряму маркетингу.

Отже, продукт є конкурентоспроможним та готовий до виходу на ринок за умови завершення прототипу.

ВИСНОВКИ ПО РОБОТІ ТА ПЕРСПЕКТИВИ ДЛЯ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Робота присвячена системному вирішенню питання впровадження маркетингових стратегій у банківських установах.

В роботі виконано огляд принципів банківського маркетингу, банківської стратегії та скорингового методу аналізу, моделювання і прогнозування кредитоспроможності клієнтів.

Розглянуто найбільш розповсюджені методи побудови скорингових моделей, а саме алгоритм логістичної регресії та дерев рішень. Виявлено особливості моделювання кредитоспроможності на основі моделей різних типів та здійснено порівняння результатів моделей на агрегованих історичних даних про кредити та клієнтів банку.

Було сформовано сучасний підхід до формування банківської стратегії для кожного цільового ринку на основі результатів скорингової моделі для оцінки кредитоспроможності потенційних клієнтів банку. Даний підхід було покладено в архітектуру застосунку системи підтримки прийняття рішень для використання у маркетинговому відділі банківської установи при впровадженні маркетингових стратегій.

Цілями даної системи підтримки прийняття рішень є:

- підвищення лояльності та довіри клієнтів банку,
- зниження витрат на маркетингові кампанії,
- зменшення кількості відмов у наданні кредитів,
- зменшення кількості заборгованості,
- збільшення операційної ефективності.

До перспектив подальшого дослідження можна віднести:

- вдосконалення розробленого методу побудови скорингової моделі;
- застосування наступних методів інтелектуального аналізу даних для підвищення точності моделей:
 - Нейронні мережі
 - Метод опорних векторів
 - Метод групового врахування аргументів
- використання даних Google Analytics та подібних сервісів для покращення маркетингової стратегії щодо пропозиції банківських продуктів.

ПЕРЕЛІК ПОСИЛАНЬ

1. Вовк, В. Я., Кредитування і контроль [Текст] : навч. посіб. / В. Я. Вовк, О. В. Хмеленко. – К.: «Знання», 2008. – 463 с. – ISBN: 978- 966-346-402-2;
2. Энциклопедия финансового риск-менеджмента / [Барбаумов В.Е., Рогов М.А., Щукин Д.Ф. и др.]; под ред. А.А. Лобанова и А.В. Чугунова. — М.: Альпина Паблишер, 2003. — 786 с.
3. Thomas L.C. Credit Scoring and its applications: Monograph / Lyn C. Thomas, David B. Edelman, Jonathan N. Crook. — Philadelphia: SIAM, 2002. — 248 p.
4. Солошенко О.М. Розробка методу k-plus-найближчих сусідів для задач машинного навчання кредитного скорингу / О.М. Солошенко // Східно-Європейський журнал передових технологій. — 2015. — Т. 3, № 9(75). — С. 29–38.
5. Boggess, W. P. Screen-test your credit risks. / W.P. Boggess // Harvard Business Review, 1967 – No 10 – pp. 113-122;
6. Witten I.H. Data mining: practical machine learning tools and techniques / Witten I.H, Eibe Frank. - Morgan Kaufmann Publishers is an imprint of Elsevier, 2011. – 664 p.
7. Кредитний ризик комерційного банку: Навч. посібник / В.В. Вітлінський, О.В. Пернарівський, Я.С. Наконечний, Г.І. Великоіваненко / За ред. В.В. Вітлінського. – К.: Т-во «Знання», КОО, 2000. – 251 с.
8. Сиддики Наим. Скоринговые карты для оценки кредитных рисков. Разработка и внедрение интеллектуальных методов кредитного скоринга / Наим Сиддики; [пер. с англ. Евгений Ильичев]. — М.: Манн, Иванов и Фербер, 2014. — 268 с.

9. В.В. Вітлінський Ризик у Менеджменті / Вітлінський В.В., Наконечний С.І. — М.; Борисфен, 1996. — 336с.

10. Siddiqi N. Credit risk scorecards: developing and implementing intelligent credit scoring / Naeem Siddiqi. — Hoboken: John Wiley & Sons, Inc., 2006. — 196 p.

11. Руководство по кредитному скорингу / [Ванг Вэй, Влатса А. Димитра, Гленнон К. Деннис и др.]; под ред. Элизабет Мэйз; [пер. с англ. И.М. Тикота; науч. ред. Д.И. Вороненко]. — Минск: Гревцов Паблишер, 2008. — 464 с.

12. Бідюк П.І. Система підтримки прийняття рішень для аналізу фінансових даних / П.І. Бідюк, Н.В. Кузнєцова, О.М. Терентьев // Наук. вісті НТУУ «КПІ». — 2011. — № 1. — С. 48–61.

13. Терентьев А.Н. SAS BASE: Основы программирования / А.Н. Терентьев, В.Н. Домрачев, Р.И. Костецкий. — К.: Эдельвейс, 2014. — 304 с.

14. Хайкин С. Нейронные сети: полный курс / Саймон Хайкин; под ред. Н.Н. Куссуль; [пер. с англ. Н.Н. Куссуль, А.Ю. Шелестова]. — 2-е изд., испр. — М.: ООО “И.Д. Вильямс”, 2006. — 1104 с.: ил.

15. Finlay S. Credit scoring, response modelling and insurance rating: a practical guide to forecasting consumer behaviour / Steven Finlay. — London: Palgrave Macmillan, 2010. — 280 p.: il.

16. Feelders A.J. Credit scoring and reject inference with mixture models / A.J. Feelders // International Journal of Intelligent Systems in Accounting, Finance and Management. — 1999. — Vol. 8, No. 4. — Pp. 271–279.